



**Maria Francisca  
Santos Eiriz**

**Caracterização molecular de estirpes autóctones de  
*Saccharomyces cerevisiae***

Dissertação apresentada à Universidade de Aveiro para cumprimento dos requisitos necessários à obtenção do grau de Mestre em Microbiologia, realizada sob a orientação científica do Professor Manuel António da Silva Santos, Professor Associado do Departamento de Biologia da Universidade de Aveiro

## **o júri**

presidente

**Prof. Dr. António Carlos Matias Correia**

professor associado com agregação Departamento de Biologia da Universidade de Aveiro

**Prof. Dr. Manuel António da Silva Santos**

professor associado do Departamento de Biologia da Universidade de Aveiro

**Prof. Dr. Laura Cristina da Silva Carreto**

professora auxiliar no Departamento de Química, Bioquímica e Ciências Farmacêuticas da Faculdade de Ciências e Tecnologia da Universidade do Algarve

**Prof. Dr. Dorit Elisabeth Schuller**

professora auxiliar do Departamento de Biologia da Universidade do Minho

## agradecimentos

À *Miss Laura*, pela robustíssima paciência com a minha impaciência e por TUDO o que me ensinou. Pelos dias desesperantes ao computador, e muito mais! À Denisa, pelos gelados, companhia fora de horas, e todas as sms. À Patrícia e Ana, pelas “que seca” juntas na salinha de tapete azul. À Cristina, pelos pães da avó com manteiga, *Ucais*. Ao João, sempre calado mas ponderado. À Tatiana, destemida “cavaleira”. À Raquel, modelo de paciência, persistência e calma. À Rita, pelas aulas de ginástica, cafés a meio da manhã e entusiasmo constante. À Isabel, pela força e ânimo, por fazer de “mãe no laboratório”. Incansáveis colegas de laboratório que à sua maneira me ensinaram ciência e vida. Pelas gargalhadas e bons momentos “escondidos”, pelo apoio, por tudo! À Catarina, pela colaboração e disponibilidade em ajudar. Ao Joel, Bruno e Miguel, que responderam sempre aos pedidos de socorro informático.

À Professora Adelaide Valente, pela ajuda com a “eterna matemática”.

Ao Professor António Correia, pela disponibilidade para resolver burocracias. Finalmente, ao Professor Manuel Santos, meu orientador, que antes de mim viu onde devia chegar e me mostrou o caminho.

À Quasi, Helena Veiga, minha “gémea” e amiga que mesmo à distância esteve sempre comigo.

Aos amigos, aqueles, os de sempre!

Aos meus pais, que mais uma vez me apoiaram e seguraram firme quando pensei não conseguir.

E ao Rui... Obr\*G@Do!

## resumo

A relação entre genótipo e fenótipo é altamente complexa. A acumulação de mutações e a adaptação dos microorganismos a diferentes condições ambientais resultam em importante diversidade genética que pode ser detectada como *single nucleotide polymorphisms* (SNPs), *indels* (deleções ou inserções), ou mesmo amplificações e alterações do número de cromossomas. A hibridação comparativa do genoma (aCGH, *array Comparative Genome Hybridization*) é uma técnica molecular de elevada aplicabilidade na identificação de variações polimórficas e na procura de diferenças genéticas que justifiquem variações fenotípicas. O objectivo deste trabalho consistiu em detectar variações genéticas, usando a técnica aCGH, entre estirpes de *Saccharomyces cerevisiae* isoladas de diferentes ambientes: estirpes autóctones de vinhas e lagares de fermentação de duas localizações geográficas distintas de Portugal, e estirpes clínicas isoladas de indivíduos imuno-deprimidos. Para tal, foram usados microarrays de DNA contendo sondas para todos os genes da levedura *S. cerevisiae* S288C, que foi usada como estirpe de referência neste estudo. A variabilidade genética entre as estirpes estudadas foi analisada, recorrendo a vários programas como CGH-Miner e SAM. Grupos de genes e variações cromossómicas característicos de determinadas estirpes foram encontrados. Os genes associados à maior variabilidade genética das estirpes estudadas pertencem às categorias funcionais de transportadores, resposta ao stress ou a drogas, factores de transcrição e tradução. Foi também possível identificar um conjunto de genes que poderão ser utilizados para criar uma futura assinatura genética das leveduras vínicas. A técnica de aCGH permitiu uma abordagem global do genoma de leveduras possibilitando diferentes análises de resultados que respondem a diferentes questões.

## **abstract**

The relationship between genotype and phenotype is highly complex. Accumulation of mutations and adaptation of microorganisms to different environmental conditions result in important genetic diversity, which can be detected as single nucleotide polymorphisms (SNPs), indels (deletions and insertions), or even amplifications or chromosome copy number changes. Comparative Genome Hybridization (aCGH) is a powerful molecular tool to identify polymorphic variation of wild type organisms and to unravel genome differences associated to phenotypic variation. In this study, we have used DNA microarrays of the complete orfeome of *Saccharomyces cerevisiae* S288C, to identify genetic variability present in wild type yeast strains and to correlate such variability with ecological adaptation. For this, yeast strains isolated from 2 different geographical locations in Portugal and also yeast clinical isolates and wine commercial strains were studied. The *S. cerevisiae* S288C strain, whose genome is completely sequenced, was used as reference for the comparative genomics study. Genome variations was analysed with CGM-Miner and SAM algorithms. Several characteristic gene groups were found within some wild-type strains, as well as unique genome alterations. Transporters, stress and drug response genes, transcription and translation factors were related to these strains' genetic variability. Part of a wine yeast signature was found and the genes were characterized. Overall, aCGH proved to be an excellent tool to access yeast genome diversity.



## ÍNDICE

Introdução .....	4
1. <i>Saccharomyces cerevisiae</i> , a levedura por excelência.....	5
2. Crescimento e ciclo de vida .....	5
3. Constituição do genoma de <i>S. cerevisiae</i> .....	7
4. Características únicas de <i>Saccharomyces sensu stricto</i> .....	10
5. Mecanismos moleculares na evolução dos genomas .....	12
5.1. Duplicação completa do genoma.....	12
5.2. Duplicação de genes e segmentos cromossómicos.....	13
5.3. Transferência lateral de genes .....	14
5.4. Duplicações subteloméricas .....	14
5.5. Aneuploidias e leveduras híbridas.....	14
5.6. Inversões, deleções e mutações pontuais.....	15
5.7. Conclusões finais.....	16
6. Genómica comparativa .....	16
6.1. Métodos comparativos.....	16
6.2. <i>Array Comparative Genome Hybridization</i> (aCGH).....	17
6.3. Aplicações de aCGH .....	19
7. Objectivos deste estudo. ....	20
Material e Métodos .....	22
1. Estirpes e condições de crescimento.....	23
2. Extracção de DNA genómico (gDNA) .....	23
3. Marcação e Hibridação .....	23
4. Aquisição e análise de dados .....	25
Resultados.....	28
1. Arrays de gDNA .....	29
2. Estirpes de <i>S. cerevisiae</i> .....	31
3. Agrupamento hierárquico das estirpes de <i>S. cerevisiae</i> .....	33
4. Cariótipos e consenso das alterações cromossómicas.....	34
5. Alterações cromossómicas de subgrupos hierárquicos de estirpes selvagens .....	44
6. <i>Significant Analysis of Microarrays</i> (SAM).....	47
7. <i>Bi-clustering</i> .....	51
8. Assinaturas de leveduras vínicas: uma comparação .....	54

Discussão geral .....	60
Conclusões.....	64
Referências .....	68
Apêndices .....	84
1. Slide de aCGH .....	85
2. Gráficos de bigodes .....	86
3. Scatter Plots .....	87
4. Cariótipos das estirpes autóctones de <i>S. cerevisiae</i> .....	89
Anexos.....	106
1. Perfis da digestão do gene <i>MET2</i> com EcoRI .....	107
2. Perfis Inter- $\delta$ .....	108
Índice de tabelas .....	110
Índice de figuras .....	112



## **INTRODUÇÃO**

## 1. *Saccharomyces cerevisiae*, a levedura por excelência

A levedura, termo que na linguagem comum se reduz a *Saccharomyces cerevisiae*, é talvez o mais antigo microrganismo domesticado. Desde a antiga civilização Egípcia até aos dias de hoje, a levedura tem sido utilizada pelo Homem. Em 1872, com os primeiros estudos de Louis Pasteur, o enorme e desconhecido potencial de *S. cerevisiae* começou a ser explorado.

Algumas das características que tornam as leveduras muito utilizadas não só em estudos científicos mas também na indústria, tecnologia e aplicações médicas ou farmacêuticas, são o seu fácil e económico cultivo com baixas necessidades nutricionais, assim como o seu estatuto de segurança (*GRAS: Generally Regarded As Safe*). Do seu estudo e aplicação surgiram vários benefícios para o Homem como a produção de proteínas heterólogas (por exemplo insulina) ou simplesmente a produção de pão, queijos ou bebidas fermentadas. Como resultado do crescente interesse e pesquisa sobre a biologia desta levedura, o seu genoma foi um dos primeiros genomas Eucariotas a ser sequenciado (Dujon, 1996; Goffeau *et al.*, 1996). A partir de então tornou-se numa referência para estudos genéticos de organismos unicelulares mas também de animais. Já na era pós-genómica *S. cerevisiae*, é utilizada como modelo para estudos de mecanismos e arquitectura celular.

## 2. Crescimento e ciclo de vida

O crescimento vegetativo destas células eucariotas unicelulares ocorre habitualmente por gemulação. Uma célula-filha cresce a partir de uma gemula da célula-mãe. Segue-se a divisão nuclear, formação da parede celular e finalmente a separação das células. Regra geral, as estirpes de laboratório têm um tempo de duplicação de 90 minutos em meio YEPD e 140 minutos em meio sintético, durante a fase exponencial à temperatura óptima de 30 °C. O tamanho e algumas características morfológicas podem variar entre estirpes e entre células haploides e diploides.

**Tabela 1. Características de células *S. cerevisiae* haploides e diploides.**

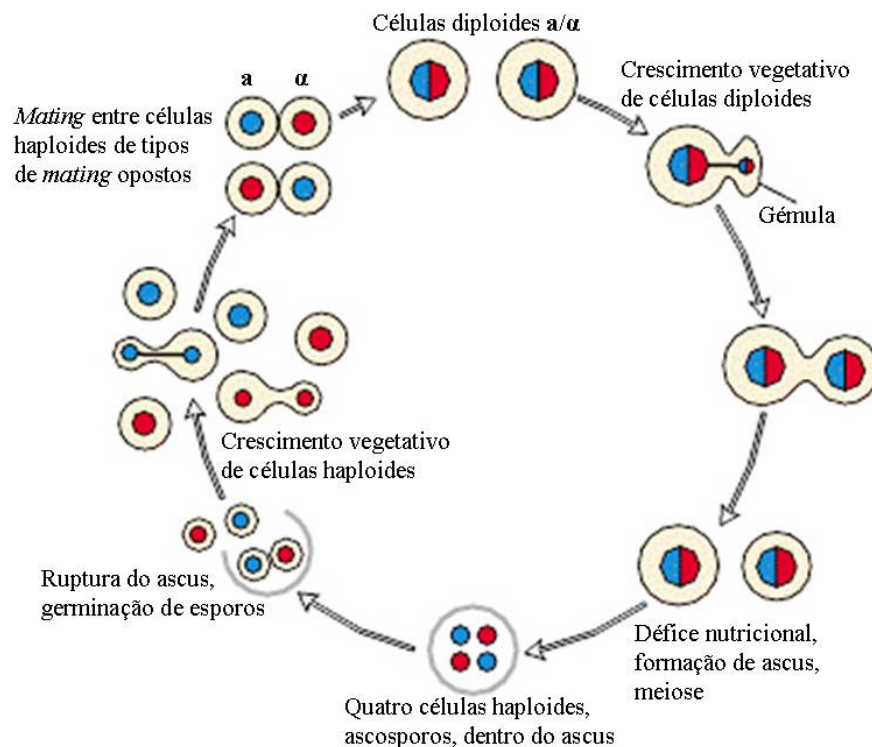
<b>Características</b>	<b>Células haploides</b>	<b>Células diploides</b>
Forma	Esférica	Elíptica
Tamanho ( $\mu\text{m}$ )	4 x 4	5 x 6
Volume ( $\mu\text{m}^3$ )	70	120
Composição ( $10^{-12}\text{g}$ ):		
Peso	60	80
Peso seco	15	20
DNA	0,017	0,034
RNA	1,2	1,9
Proteínas	6	8

Ao contrário de muitos outros microrganismos, o genoma de *S. cerevisiae* é estável tanto na forma haploide como diploide. Esta conversão entre estados de haploidia e diploidia é conseguida pelo equilíbrio entre esporulação e crescimento mitótico. A capacidade de duas células haploides **a** e  **$\alpha$**  se conjugarem, fundirem e originarem uma célula diploide **a/ $\alpha$**  é denominada de *mating* ou reprodução sexuada. O tipo de *mating* de cada célula é determinado pelo *locus MAT* (**MATing type**).

**Tabela 2. Características de *mating* dos diferentes tipos de células.**

	<b>MATa</b>	<b>MAT<math>\alpha</math></b>	<b>MATa/MAT<math>\alpha</math></b>
Tipo de célula	<b>a</b>	<b><math>\alpha</math></b>	<b>a/<math>\alpha</math></b>
Reprodução sexuada	SIM	SIM	NÃO
Capacidade de esporulação	NÃO	NÃO	SIM
Feromona	Factor <b>a</b>	Factor <b><math>\alpha</math></b>	NENHUMA
Receptor	Liga factor <b><math>\alpha</math></b>	Liga factor <b>a</b>	NENHUM

Células com o alelo **MAT $\alpha$**  são do tipo  **$\alpha$** , secretam pequenos péptidos (feromonas) deste tipo e expressam na superfície celular receptores de péptidos de tipo complementar. Similarmente, células com o alelo **MATa** são do tipo **a**, secretam feromonas **a** e reconhecem feromonas  **$\alpha$**  à sua superfície. O reconhecimento de feromonas de tipo complementar leva à aproximação das células haploides em questão, fusão celular e nuclear numa única célula diploide heterozigótica **a/ $\alpha$** . Estas células podem, mediante as condições ambientais, dividir-se mitoticamente ou, se sob défice nutricional, dividir-se meioticamente. Neste caso originarão 4 células haploides, duas de cada tipo, **a** e  **$\alpha$** . Se as condições o proporcionarem, os esporos germinarão e dividir-se-ão mitoticamente (Lewin 2004).



**Figura 1. Representação esquemática do ciclo de vida de *S. cerevisiae*.**

Duas células haploides de diferentes tipos de mating, **a** e **α**, podem unir-se originando uma célula diploide **a/α** que se multiplica por gemulação. Sob défice nutricional, as células diploides entram em meiose formando ascósporos. A ruptura de ascos liberta quatro esporos que germinam como células haploides. A cada nova geração, uma célula haploide pode trocar de tipo (Adaptado de Lodish et al., 2001).

Algumas espécies de leveduras são capazes de trocar de tipo sexual ao longo das suas gerações (ver página 12). Esta capacidade está definida numa *casset* presente no cromossoma III. O locus central, *MAT*, é transcrito enquanto os locus *HML* (*Hidden Mat Left*) e *HMR* (*Hidden Mat Right*) que o flanqueiam estão silenciosos. A cada nova geração, a sequência de um destes *locus* pode ser transferida para o *locus* central, que sendo expresso, promove a troca de tipo de *mating*. O alelo *HML* propicia troca de **α** para **a** e o *HMR* de **a** para **α**.

### 3. Constituição do genoma de *S. cerevisiae*

Em 1989, 35 laboratórios Europeus criaram um consórcio para sequenciar o genoma da levedura (Goffeau *et al.*, 1991). O cromossoma III foi o primeiro a ser sequenciado (Oliver *et al.*, 1992) e em Abril de 1996 todo o genoma estava sequenciado e o projecto terminado.

*The Yeast Genome Directory* (Goffeau *et al.*, 1997) foi publicado um ano depois com o mapa genético completo de *S. cerevisiae* S288C e um inventário com a classificação de todas as proteínas de levedura e as estratégias de sequenciação aplicadas. Os 12,8 Mb de DNA do genoma estão organizados em 16 cromossomas e alguns plasmídeos. Contém aproximadamente 5800 ORFs (**O**pen **R**eadin**G** **F**rames) que codificam proteínas específicas, com espaçamento médio de 2 kb (Dujon, 1996). Esta densidade de genes é 200 vezes superior à do genoma Humano, mas apenas 4 vezes superior à de *Escherichia coli*.

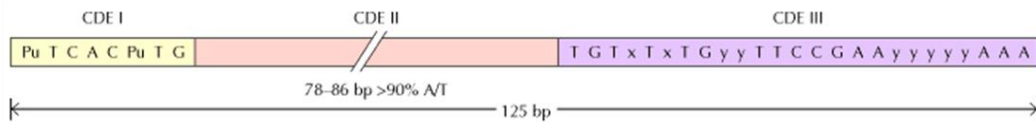
**Tabela 3. Características fundamentais do genoma de leveduras.**

(Adaptado de Broach J. et al., *The molecular and cellular biology of the yeast S. cerevisiae*, 1991)

Tamanho do genoma	12,8 Mb + rDNA
ssDNA	12,0 Mb
Comprimento mapa genético	> 4300 cM
Tamanho dos cromossomas	250 a 2500 kb
DNA em organelos	Mitocondrial (75 kb de baixa densidade e rico em A e T)
Elementos extracromossómicos	Plasmídeo circular de 2 µm + dsRNA

O DNA ribossomal (rDNA) é codificado por genes (rDNA genes) que existem em aproximadamente 120 cópias no cromossoma XII. Cada grupo de rDNAs é composto por unidades de transcrição alternadas com intrões. Os tRNA são codificados por um grande número de cópias de genes presentes ao longo do genoma (Mortimer, 1985). Existem, ainda RNAs de cadeia dupla (dsRNAs) que codificam toxinas e componentes fundamentais à transcrição e replicação de vírus de leveduras. O DNA mitocondrial, assim como os elementos extracromossómicos, não são fundamentais para a sobrevivência da levedura, apenas são necessários à sua própria replicação e continuidade. O mtDNA de *S. cerevisiae* é cinco vezes mais longo que o do Homem pois existem longos intrões responsáveis pelo aumento do tamanho das grelhas de leitura (Open Reading Frames ou ORFs).

Muitas estruturas do genoma têm importantes funções reguladoras a nível cromossómico. Por exemplo, as sequências *cis* [Autonomous Replication Sequences (ARS), Centrómeros (CEN) e telómeros (TEL)]. As ARSs são pequenas sequências ricas em A·T que asseguram capacidade de replicação autónoma. Têm um domínio A com 14 bp das quais 11 são conservadas. Os elementos CEN são sequências de aproximadamente 125 bp responsáveis pela correcta segregação cromossómica durante mitose e meiose. Estes elementos possuem 3 tipos de sequências: CDE-I, II e III.



**Figura 2. Constituição dos centrômeros de *S. cerevisiae*.**

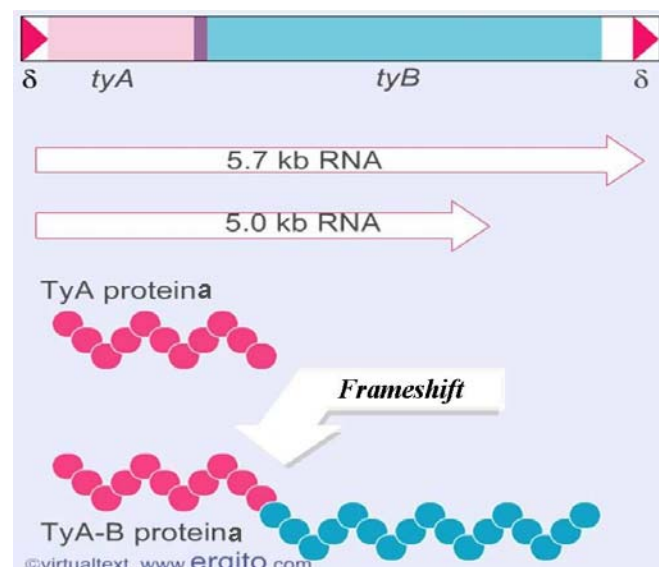
As sequências conservadas CDE-I e CDE-III estão separadas por 78 a 86 bp de DNA rico em A·T (CDE-II). As sequências apresentadas são o consenso da análise de centrômeros de cromossomas individuais de leveduras. Pu = A ou G; x = A ou T; y = qualquer base (Adaptado de Cooper, 2000).

Para além das referidas sequências *cis*, os telômeros, as repetições subteloméricas e os retrotransposições ou elementos *Ty* são importantes na caracterização do genoma de leveduras. Telômeros são sequências repetidas de  $(TG_{1-3})_n$  de tamanho variável mas rondando os 300 nucleótidos (Shampay *et al.*, 1984). Estas repetições impedem o encurtamento dos cromossomas a cada novo ciclo de replicação celular uma vez que a telomerase os alonga de forma análoga à transcriptase reversa (Cech *et al.*, 1997). Associadas aos telômeros aparecem as repetições subteloméricas (Chan & Tye, 1983). Aparentemente sem função cromossômica, estas sequências podem ser de duas classes, X ou Y', variando o seu tamanho e número de estirpe para estirpe (Louis & Harber, 1990). X são sequências menos conservadas em tamanho e localizadas no centro das sequências Y'. Y' são sequências adjacentes aos telômeros variando bastante no número de cópias.

Os elementos *Ty*, (yeast *transposons*), representam uma família de sequências repetitivas de DNA. Os retrotransposições, com organização semelhante aos retro vírus, são LTRs (*Long Terminal Repeat*) de 5 a 6 kb dispersos por todo genoma sem capacidade própria de se moverem. A transcriptase reversa, gerando intermediários de RNA com capacidade de transposição, permite a deslocação destes fragmentos pelo genoma (Roeder & Fink, 1982). A sua taxa de transposição é inferior à habitualmente observada na maioria dos transpões ( $\sim 10^{-7}$  a  $10^{-8}$ ) devido ao balanço constante entre transposição e excisão (Fink *et al.*, 1981). Pelo facto de se moverem com facilidade e recombinarem de diversos modos, estes elementos têm grande influência na evolução e modelação do genoma de leveduras.

Existem vários tipos de elementos *Ty*, nomeadamente *Ty1* a *Ty5*, representando duas grandes categorias: *Ty1-copia* (*Ty1*, *Ty2*, *Ty4* e *Ty5*) e *Ty3-gypsy* (apenas *Ty3*) (Kim *et al.*, 2007). Embora pertencentes a diferentes categorias, todos os elementos *Ty* partilham a mesma organização e codificam as mesmas proteínas. Estes retrotransposições possuem duas

ORFs com 13 nucleótidos sobrepostos que são expressas na mesma direcção mas em diferentes fases. Um mecanismo de *frameshifting* é utilizado para ultrapassar o codão STOP no final da ORF *TyA* e expressar *TyB*. Enquanto *TyA* é expressa como todas as outras proteínas, através de uma ORF que codifica proteínas estruturais da partícula tipo-vírus; *TyB* é expressa como uma proteína quimérica, contendo uma protease, uma integrase, uma transcriptase reversa e o domínio catalítico da ribonuclease H (Lesage & Todeschini, 2005).



**Figura 3. Representação esquemática de mecanismo de *frameshift* responsável pelas duas proteínas produzidas a partir dos elementos *Ty*.**  
(Adaptado de [www.ergito.com](http://www.ergito.com)).

#### 4. Características únicas de *Saccharomyces sensu stricto*

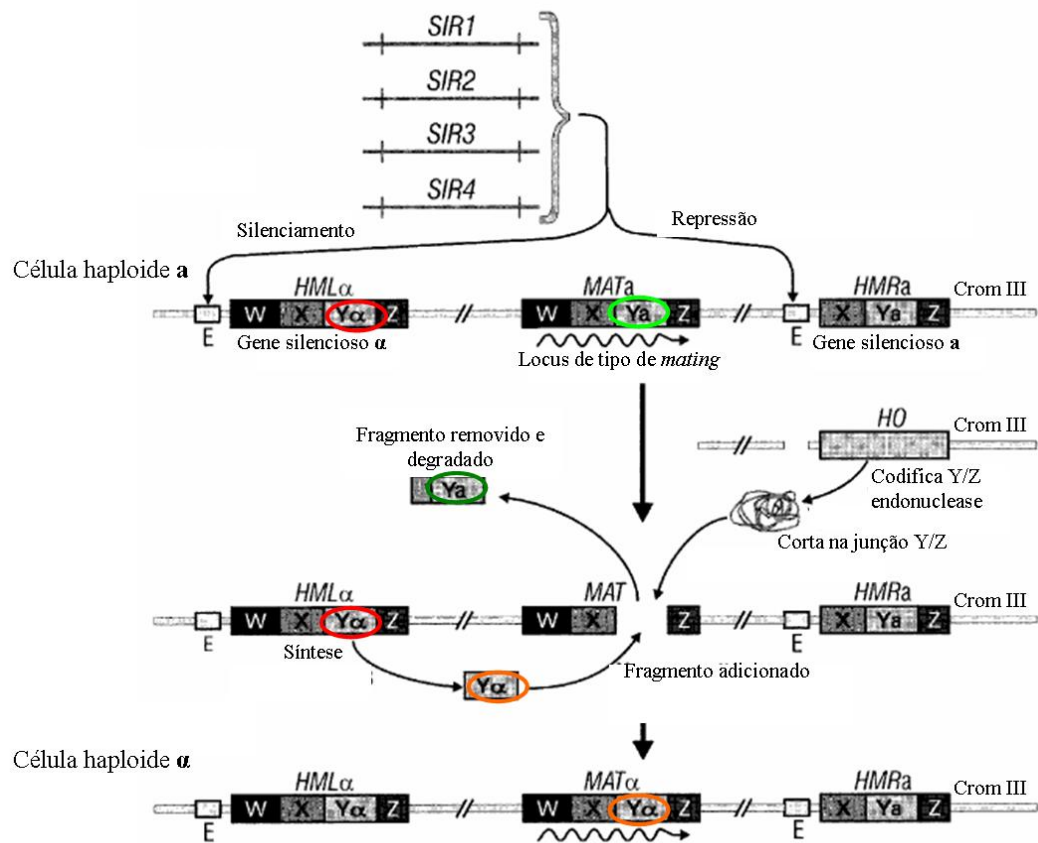
*S. cerevisiae*, *S. bayanus*, *S. paradoxus*, *S. cariocanus*, *S. kudriavzevii* e *S. mikatae*, são as leveduras do género *Saccharomyces* que constituem um grupo denominado *Saccharomyces sensu stricto*. Todas estas leveduras apresentam elevadas semelhanças fenotípicas sendo a sua distinção por vezes árdua. Este grupo de leveduras possui algumas características fenotípicas comuns como a gemulação multilateral, a facilidade de esporulação mesmo sob condições nutricionais favoráveis, sendo as células destas leveduras habitualmente diploides (Rainieri *et al.*, 2003). Algumas destas características têm a sua origem em diferenças genotípicas adquiridas ao longo da evolução, como por

exemplo o mecanismo de repressão do metabolismo da glucose (*glucose repression*), características de *mating* próprias e a enzima dihidroorotato desidrogenase (DHODase).

Estas leveduras conseguem sobreviver na ausência de oxigénio (Møller *et al.*, 2001), mas também podem fermentar na presença deste. *Glucose repression* é um mecanismo de inibição do metabolismo de outros açúcares que não a glucose, desencadeado por este último. Várias são as proteínas envolvidas neste processo, Mig1p, CREAp, Tup1p, Ssn6p, Snf1p, Hxk2p, Gal4p, Hap4p e Ume5 são alguns exemplos. As enzimas responsáveis por esta preferência pela glucose são reguladas não só ao nível pós-tradução mas também ao nível da sua transcrição. Na ausência de oxigénio estas células são capazes de fermentar a glucose produzindo etanol. Esta conversão de glucose em etanol é uma vantagem evolutiva, uma vez que muitos microorganismos não sobrevivem na presença deste álcool.

Tal como foi acima referido, algumas leveduras são capazes de trocar de tipo sexual (*mating type*). Para além dos elementos silenciosos da *casset* que regula o *mating* de cada célula, genes como *STE* (*STERile*) e *HO* (*HO*mothallic switching endonuclease) são fundamentais neste processo. A eliminação da capacidade de *mating* é regulada pelo gene *STE* e o favorecimento desta troca é regulado pelo gene *HO* (Herskowitz, 1992). O alelo dominante *HO* aumenta a capacidade de troca de tipo sexual a cada nova geração. A endonuclease *HO* cliva o locus *MAT* permitindo a sua substituição por um dos alelos silenciosos. Por sua vez, o alelo recessivo *ho* diminui esta taxa para aproximadamente  $10^{-6}$ . Uma vez que o *Saccharomyces sensu stricto* possui o alelo *HO*, estas espécies podem de trocar de tipo **a** ou  **$\alpha$** .





**Figura 4. Modelo da *casset* responsável pela troca de tipo sexual (*mating type*).**  
(Adaptado de Pretorius & Van der Westhuizen, 1991)

DHODase é uma enzima mitocondrial codificada pelo gene *URA9* (*URAcil requiring*), que converte dihidroorotato em orotato no quarto passo da biossíntese de pirimidinas, (Jensen *et al.*, 1998; Nagy *et al.*, 1992). A DHODase necessita, na presença de oxigénio, de mitocôndrias funcionais para que a quinona funcione como receptor final de electrões na biossíntese de uracilo. Em *Saccharomyces sensu stricto* a DHODase é citosólica sendo codificada pelo gene *URAI*. Neste caso, DHODase citosólica, na ausência de oxigénio, há biossíntese de uracilo pois o fumarato citosólico passa a ser o receptor final dos electrões (Gojkovic *et al.*, 2004).

## 5. Mecanismos moleculares na evolução dos genomas

### 5.1. Duplicação completa do genoma

A duplicação completa do genoma, WGD (*Whole Genome Duplication*), ocorreu em vários organismos incluindo em *S. cerevisiae*. Algumas características do genoma desta levedura comprovam esta duplicação, nomeadamente a presença de regiões homólogas conservadas em alguns cromossomas (CHRs, **C**onserved **H**omologous **R**egions) (Wolfe & Shields, 1997), famílias de genes representadas por várias cópias de genes não totalmente idênticos dispersos pelo genoma e regiões subteloméricas. Esta duplicação e consequente formação de diferentes genomas pela acumulação de eventos moleculares, pode sintetizar-se da seguinte forma. Um genoma ancestral dá origem a duas linhas celulares distintas. Uma é idêntica à ancestral e por isso estável, outra sofre uma duplicação de todo o genoma e é instável. Este genoma instável, com cópias parálogas de todos os genes (genes homólogos gerados por duplicação) vai sofrer alterações. *Gross Chromosomal Rearrangements* (GCRs) são rearranjos genéticos: inversões, translocações e também deleções de uma ou várias cópias de genes. A acumulação de mutações, levando à especialização de determinados genes, evolução de determinadas funções ou divergência de sequências podem também ocorrer (Liti & Louis, 2005). Genes parálogos resultantes de WGD têm maior probabilidade de interação e partilha de funções biológicas do que genes parálogos resultantes de duplicações em menor escala. Guan e colaboradores (2007) sugerem que genes parálogos resultantes de WGD conferem vantagem evolutiva sob diferentes condições ambientais.

## 5.2. Duplicação de genes e segmentos cromossómicos

A duplicação de genes ou fragmentos cromossómicos em genomas eucariotas é frequente. A maioria destes eventos ocorre no mesmo cromossoma (intra-cromossómicas) em fragmentos de 40 a 650 Kb. As duplicações inter-cromossómicas são menos frequentes. Koszul e colaboradores (2004) demonstraram que células haploides de *S. cerevisiae* em mitose duplicam espontaneamente fragmentos cromossómicos a uma taxa de  $\sim 10^{-9}$  a  $10^{-10}$ . Para evitar a expansão descontrolada do genoma, a perda de genes funciona como balanceador das duplicações. Este equilíbrio é conseguido por deleções mas mais frequentemente pela inativação de genes duplicados. Esta ocorre pela acumulação de mutações *nonsense* gerando pseudo genes. Assim, duplicações de fragmentos cromossómicos que não sejam vantajosas tendem a desaparecer ao longo da evolução do

genoma pois a sobrevivência e funcionalidade da célula depende apenas de uma única cópia dos genes em causa (Brookfield, 1992).

Exemplos de sucesso de duplicação de segmentos cromossômicos com vantagens para as células são o caso da aquisição de anaerobiose facultativa das espécies do *Saccharomyces sensu stricto*, assim como do alelo dominante *HO* que lhes possibilita a troca de tipo sexual (*mating type*) (Piskur & Lagkjær, 2004). Verifica-se por vezes, a existência de caudas poli(A) no final de fragmentos duplicados inter- e intra-cromossômicos indicando que intermediários de RNA estiveram na origem de algumas das referidas duplicações (Liti & Louis, 2005).

### 5.3. Transferência lateral de genes

Tanto os intrões, como a transferência horizontal de genes, são pouco frequentes em leveduras (Langride, *et al.*, 1991; Dujon, *et al.*, 2004). Esta última representa menos de 0,2 % do material genético. Quando identificados, genes específicos de espécies estão associados a funções enzimáticas podendo conferir vantagens evolutivas em determinados nichos ecológicos. Por exemplo, a independência de oxigénio na biossíntese de pirimidinas das leveduras de *Saccharomyces sensu stricto* parece ter sido obtida pela transferência horizontal de genes de *Lactococcus lactis* (Nagy *et al.*, 1992; Gojkovick *et al.*, 2004; Hall *et al.*, 2005).

### 5.4. Duplicações subteloméricas

A recombinação ectópica entre elementos Y' permite a amplificação e diversificação da sequência das regiões subteloméricas (Louis & Harber, 1992). A estrutura destas regiões, sendo semelhante entre os eucariotas, revela um possível papel na evolução adaptativa de genomas. Muitos dos genes aqui presentes, embora não fundamentais para a célula, desempenham funções importantes em novas condições ambientais por estarem relacionados com metabolismo secundário (Pryde *et al.*, 1997).

### 5.5. Aneuploidias e leveduras híbridas

À medida que a ploidia das células aumenta, a perda de um cromossoma afecta o fenótipo desta mas não a inviabiliza, como aconteceria se esta fosse haploide. Assim, em células poliploides, um número ímpar de cromossomas, por perda ou ganho destes, é denominado aneuploidia e representa um outro mecanismo de evolução. Esta aneuploidia é adquirida devido a uma incorrecta segregação cromossómica durante a meiose (Liti & Louis, 2005).

A formação de híbridos inter-específicos constitui também uma forma de evolução e especiação. Esporos ou células haploides de diferentes espécies do género *Saccharomyces*, mais precisamente de *Saccharomyces sensu stricto*, podem conjugar-se formando híbridos (Greig *et.al.*, 2002). Estas espécies conseguem efectuar *mating* entre si mas os seus híbridos são meioticamente estéreis. Embora se forme um zigoto híbrido, raramente os esporos resultantes são viáveis. Esta infertilidade parece estar associada à perda diferencial de genes que segue a WGD. Muito esporadicamente ocorre a duplicação do genoma destes híbridos que assim têm 32 cromossomas complementares capazes de recombinação (Liti & Louis, 2005). Caso não ocorra a dita duplicação, a fertilidade destes híbridos apenas se restabeleceria se a possibilidade de recombinação fosse eliminada (Dujon, 2006). Assim, embora a barreira pós-zigótica já exista no *Saccharomyces sensu stricto*, a barreira pré-zigótica inexistente indica uma fase inicial de especiação (Piskur & Langkjaer, 2004).

#### 5.6. Inversões, deleções e mutações pontuais

Quando se encontra um gene invertido, a presença desta inversão pode justificar-se pela deleção da cópia original do gene depois de uma duplicação invertida (Fisher *et al.*, 2001). Mutações simples num único nucleótido de um gene podem não ter qualquer efeito detectável no fenótipo de uma espécie ou estirpe ou, pelo contrário, causarem grande alteração. O impacto destas mutações na expressão de um gene depende de alguns factores, entre os quais o número de outros genes que influenciam a expressão do gene em questão (*trans-mutational target size*) e o número de elementos reguladores do gene em questão (*cis-mutational target size*). Genes com mais locais de ligação a factores de transcrição (*cis-mutational target size*) são mais susceptíveis aos efeitos de mutações espontâneas do que outros (Landry *et al.*, 2007). Assim, estirpes com combinações precisas de genes com mutações pontuais podem ter fenótipos bastante precisos e específicos.

## 5.7. Conclusões finais

Os diversos mecanismos moleculares apresentados mostram que diferentes espécies, ou mesmo estirpes de leveduras, poderão ter diferentes conteúdos genéticos gerados pela combinação dos referidos mecanismos moleculares. A duplicação de genes ou de todo genoma e a perda de destes estabelece um equilíbrio dinâmico entre formação e perda de genes. A manutenção ou eliminação de mutações pontuais, intrões, transposões, retrotransposões e transferência horizontal têm e continuarão a ter um papel fundamental na composição do material genético de cada estirpe e de cada espécie.

## 6. **Genómica comparativa**

A comparação esteve sempre adjacente à evolução da biologia. Prova disso foi o impacto dos estudos de Darwin no pensamento científico. A genómica comparativa, um dos mais recentes campos da ciência, permite a análise e comparação de genomas de diferentes espécies. Uma melhor compreensão da evolução das espécies e dos seus genomas, bem como de cada uma em particular, é o objectivo último desta ciência. Muito deste avanço dependeu, e ainda depende, da sequenciação de um grande número de genomas. Técnicas computacionais e laboratoriais, em conjunto, são fundamentais ao avanço do conhecimento desta área.

### 6.1. Métodos comparativos

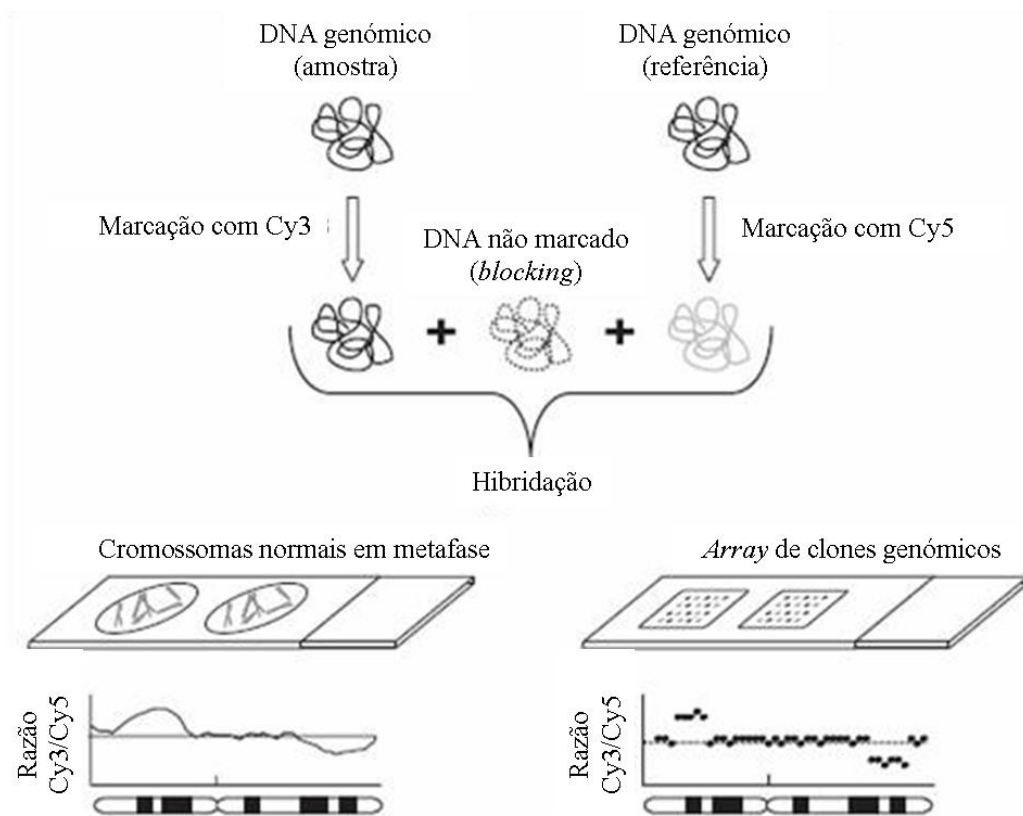
Vários métodos são usados para estudar comparativamente diferentes espécies ou estirpes da mesma espécie, desde a comparação do tamanho ou densidade dos genomas, passando pela comparação dos proteomas não-redundantes ou a anotação estrutural e funcional dos genomas. Dos métodos computacionais destacam-se os alinhamentos e as reconstruções filogenéticas. Através dos alinhamentos múltiplos de sequências, organizações de vários genomas ou proteomas, possíveis elementos regulatórios ou presumíveis funções de regiões codificantes sem anotações são identificadas. Quando tais alinhamentos são efectuados, é visualizada a história evolutiva. Assim, as reconstruções filogenéticas não são mais do que árvores que demonstram as relações evolutivas entre os genomas alinhados (Haubold & Wiehe, 2004).

Ao nível laboratorial, a sequenciação completa do genoma de diferentes espécies e o constante avanço de técnicas de análise de genomas completos tem sido fundamental. A contribuição dos *microarrays* nesta análise tem sido fulcral. Variações na sequência genómica de diferentes espécies ou estirpes podem ser identificadas utilizando *microarrays*. Assim, o reconhecimento das diferenças genéticas responsáveis pelas diferenças fenotípicas como fim último pode ser atingido (Fay *et al.*, 2004; Steinmetz *et al.*, 2002).

## 6.2. Array Comparative Genome Hybridization (aCGH).

O primeiro estudo de CGH foi descrito em 1992 (Kallioniemi *et al.*) com DNA genómico diferencialmente marcado e hibridado com cromossomas em metafase, desnaturados e imobilizados num slide. No entanto, a baixa precisão na detecção de pequenas aberrações estava limitada à resolução dos cromossomas em metáfase como mapa. Foi então que surgiram os primeiros *arrays* para realização de CGH (Solinas-Toldo *et al.*, 1997; Pinkel *et al.*, 1998; Pollack *et al.*, 1999). O aCGH permitiu ultrapassar o problema na precisão da detecção de pequenas alterações cromossómicas. Em 1999, 3360 sondas de cDNAs humanos imobilizados num chip foram utilizadas para detectar perdas e ganhos de DNA no cancro da mama (Pollack *et al.*, 1999).

O mapa das variações do número de cópias de DNA de uma amostra relativamente a uma referência é feito de forma qualitativa e quantitativa. A razão da ligação de DNA marcado às sequências complementares impressas no *array* é proporcional à concentração relativa de DNA na amostra inicial. Assim, cálculos quantitativos da hibridação ao longo do mapa genético fornecem informação sobre o número de cópias relativas de diferentes sequências do genoma. Este método baseia-se na marcação diferencial de duas amostras e hibridação competitiva num slide com uma matriz de sondas imobilizadas, *array*. Habitualmente o DNA da amostra em estudo é marcado com uma molécula fluorescente (Cy3 ou análoga) e o DNA de referência é marcado com outra molécula fluorescente (Cy5 ou análoga). Ambos são simultaneamente hibridados num slide. A leitura óptica é feita em cada um dos canais correspondentes aos fluoróforos utilizados. Depois, através da utilização de software adequado, filtração e análise de dados, assim como controlo de qualidade de toda a hibridação, é possível obter informação de alta qualidade sobre variações genómicas.



**Figura 5. Comparação entre CGH e aCGH.**

DNA genômico da amostra é marcado com um fluoróforo (p. ex. Cy3) e o DNA de referência é marcado com outro fluoróforo (p. ex. Cy5). Ambos DNAs são misturados com DNA não marcado para bloquear sequências repetitivas do genoma (blocking). Após desnaturação, esta mistura é hibridada com cromossomos em metáfase (no caso de CGH) ou com uma matriz de clones genômicos (para o aCGH). Depois da hibridação, a razão entre o sinal dos dois fluoróforos é medida ao longo dos cromossomos imobilizados ou em cada clone de DNA genômico (Adaptado de Snijders et al. 2003).

A metodologia inicial neste tipo de técnica centra-se na construção de um *array* genômico. As sondas utilizadas podem ter diferentes origens. Existem BAC *arrays*, *arrays* de cDNA e *arrays* com sondas de produtos de PCR.

**Tabela 4. Resumo de algumas características dos diferentes tipos de arrays impressos para aCGH.**

<b>Plataforma</b>	<b>Sondas</b>	<b>Descrição</b>	<b>Desvantagens</b>
Grandes fragmentos de clones genómicos	BACs	DNA de clones isolado impresso nos arrays	Insuficiente quantidade de DNA e com peso molecular não adequado a uma boa impressão - elevada viscosidade (Solinas-Toldo, <i>et al.</i> , 1997; Pinkel, <i>et al.</i> , 1998; Albertson, <i>et al.</i> , 2000)
Fragmentos de PCR	Fragmentos de PCR	Representantes de clones impressos nos arrays	Elevados custos e muito trabalhosos (Barrett, <i>et al.</i> , 2004).
Arrays de cDNA	cDNA, EST	cDNAs impressos nos arrays (bibliotecas disponíveis)	Não correspondem ao genoma completo e muitas ESTs não estão anotadas (Pollack, <i>et al.</i> , 1999)

Os *arrays* de cDNA, descritos pela primeira vez por Schena *et al.* (1995), são habitualmente usados em estudos de expressão genética. Mais tarde foi-lhes também atribuída a função complementar no estudo da variação do número de cópias de genes de todo o genoma (Pollack *et al.*, 1999). Para além da facilidade de obtenção deste tipo de sondas, é possível realizar estudos de expressão genética com os mesmos *arrays* com que se analisam os números de cópias de DNA de um genoma. Das desvantagens destes *arrays* constam a limitação aos genes conhecidos e anotados, a impossibilidade de estudar intrões, promotores ou sequências intra-génicas. A redundância de algumas sequências de genes parálogos dificulta também a análise dos dados de aCGH. Todos os arrays têm uma desvantagem em comum, a impossibilidade de detectar rearranjos equilibrados no genoma sem alteração do número de cópias de DNA.

### 6.3. Aplicações de aCGH

Deleções, amplificações e translocações são, como já foi referido, alguns dos mecanismos que podem estar subjacentes aos desarranjos e alterações genéticas. A análise de alterações cromossómicas ao nível de amplificações de oncogenes ou deleções de genes supressores de tumores é particularmente importante no estudo do desenvolvimento de tumores, sua detecção e tratamento (Kinzler & Vogelstein, 1996) e tem sido um dos campos mais explorados com aCGH.



Estudos filogenéticos para comparação das diferenças genéticas entre isolados da mesma espécie ou de espécies presumivelmente semelhantes ou evolutivamente próximas, podem ser efectuados com aCGH (Lashkari *et al.*, 1997; Edwards-Ingram *et al.*, 2004). Neste campo, dependendo do tipo de arrays, diferentes conclusões podem ser retiradas. Arrays que contêm todo o genoma, regiões codificantes e não codificantes, são mais informativos do que os arrays que apenas contêm regiões codificantes. No primeiro caso, alterações em todo o genoma são detectadas. No segundo, apenas alterações em fragmentos correspondentes a genes (ORFs) ou blocos de genes são encontradas. Há ainda arrays que permitem detectar pequenas alterações ao nível de um único nucleótido, os SNP - *single nucleotide polymorphisms*.

## **7. Objectivos deste estudo.**

Diferentes estirpes de leveduras, provenientes de diferentes nichos ecológicos, têm diferentes fenótipos. Alguns estudos demonstraram a existência de polimorfismos em estirpes selvagens de leveduras vínicas e a correlação entre genótipo e fenótipo (Nadal *et al.*, 1996; Esteve-Zarzoso *et al.*, 2004). Neste estudo tentámos explorar em detalhe esta relação entre genótipo e fenótipo usando aCGH. Para tal, estudámos variações genotípicas em estirpes selvagens de *S. cerevisiae* isoladas de vinhas e lagares da Bairrada, estirpes seleccionadas em ambiente industrial, estirpes selvagens isoladas na Região do Vinho Verde e também estirpes patogénicas isoladas na clínica. Considerando as grandes diferenças ecológicas entre as estirpes ambientais, as seleccionadas na indústria e os isolados clínicos, pareceu-nos que tais estirpes seriam um excelente ponto de partida para avaliar a influência do ambiente na estrutura primária do genoma e clarificar a relação entre genótipo e fenótipo.



## **MATERIAL E MÉTODOS**

## **1. Estirpes e condições de crescimento**

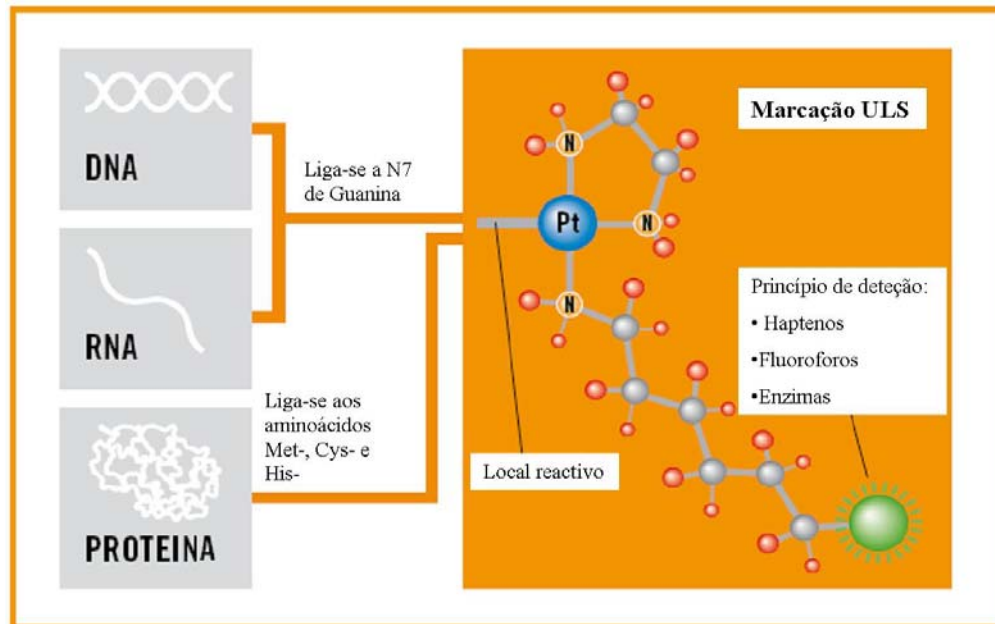
A lista que se segue discrimina as 15 estirpes selvagens de *S. cerevisiae* de diferentes ambientes utilizadas neste estudo: Davis Lalvin 522, IOC 18-2007, AEB Fermol Rouge, ICV D245, 06L3FF02, 06L1FF11, 06L3FF15, 06L6FF20, BB 1235, BB 2453, UM 218, UM 237, J940047, J940557 e J940915. Uma estirpe adicional foi utilizada como referência universal: *S. cerevisiae* S288C (MATa SUC2 *mal mel gal2 CUP1 flo1 flo8-1 hap1*). As células foram cultivadas em meio líquido YPD (1 % extracto levedura, 2 % peptona, 2 % glucose), durante a noite ( $\sim 10^8$ - $10^9$  células), a 30 °C e com agitação constante, 185 rpm.

## **2. Extracção de DNA genómico (gDNA)**

O protocolo de extracção de DNA genómico foi adaptado de Hoffman & Winston (1987) e optimizado para minimizar contaminação do DNA com outros ácidos nucleicos ou proteínas. Culturas de 5 ml em YEPD, foram recolhida e lavadas 3 vezes com água ultra pura. Após centrifugação, as células foram ressuspensas em 200 µl tampão de lise (2 % Triton X-100, 1 % SDS, 100 mM NaCl, 10 mM Tris pH 8.0, 1mM EDTA pH 8.0). Foram adicionados 200 µl de 25:24:1 fenol:clorofórmio:álcool isoamílico e 300 mg de esferas de vidro lavadas em ácido (diâmetro  $\sim 0,5$  mm). A mistura foi agitada num vortex durante 10 minutos. Adicionaram-se 200 µl TE (10 mM Tris-HCl, pH 8.0, 1 mM EDTA, pH 8.0) e após 5 minutos de centrifugação à velocidade máxima o DNA presente na fase aquosa foi transferido para novo eppendorf onde foi precipitado com 1 ml etanol 100% à temperatura ambiente. Para remoção completa do RNA, o gDNA foi ressuspenso em 400 µl TE com 60 mg RNaseA e incubado 6 horas a 37 °C. Seguidamente, para remover as proteínas, incubou-se o gDNA, durante a noite, a 50 °C com 10 mg Proteinase K. Finalmente, o gDNA foi precipitado com 20 µl Acetato Sódio 3.0 M (pH 5.2) e 500 µl de etanol 100% frio, a -80 °C durante 1 hora. Depois de seco, o gDNA foi ressuspenso em 200 µl TE.

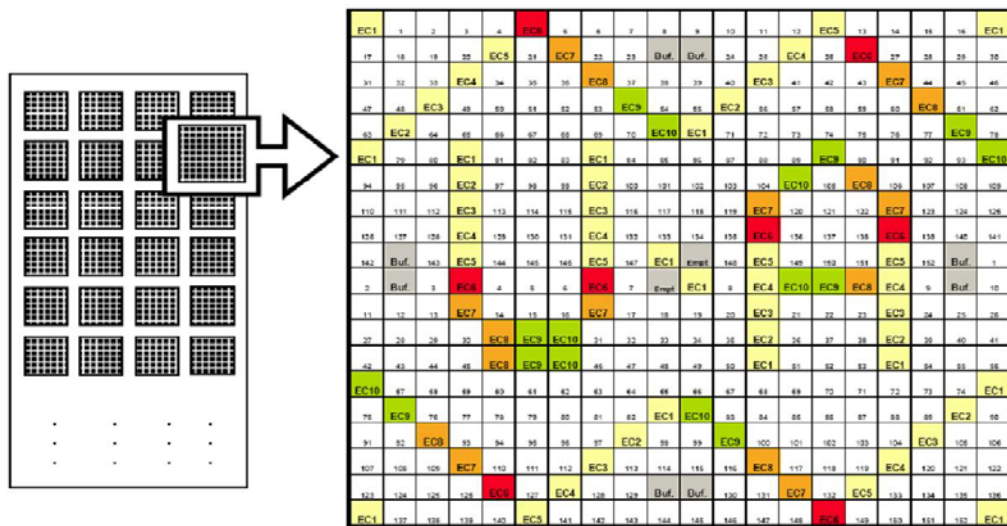
## **3. Marcação e Hibridação**

Depois de digerido com DpnII (fragmentos entre 250 bp to 3000 bp), o DNA foi marcado com fluorescência usando o sistema ULS da Kretech (Genome-pULSe ULS arrayCGH Whole Genome Labelling Kit). Como se pode observar na figura 6, é absolutamente necessário que todo o gDNA esteja limpo e sem RNA ou proteínas contaminantes.



**Figura 6. Sistema de marcação fluorescente ULS pela formação de uma ligação coordenada através do Azoto (N7) da Guanina.**

Em todas as experiências de aCGH foram utilizados slides de microarrays com sondas de 70 nucleótidos correspondentes ao conjunto de todas as ORFs de *S. cerevisiae* S288C (AROS v1.1, OPERON, Quiagen). Todas as sondas foram impressas em duplicado. Sondas de DNA de *E. coli* foram utilizadas como controlos negativos (quadrículas coloridas na figura 7).



**Figura 7. Esquema da matriz impressa nos slides utilizados em aCGH.**

As quadrículas coloridas representam controlos negativos para aCGH sendo também úteis no controlo de qualidade de experiências de expressão genética. Cada slide tem 4 x 20 blocos cada um com 20 x 20 sondas de cDNA (Figura gentilmente cedida por Laura Carreto).

Antes de cada hibridação os slides foram bloqueados durante 45 minutos a 42 °C com uma solução filtrada (0,22 µm) de 5x SSC (0,15 M Cloreto de sódio, 0,015 M Citrato de sódio, pH 7,0), 25% Formamida, 0,1% SDS e 1% BSA. Após esta pré-hibridação os slides foram lavados 5 vezes com água ultra pura e 5 vezes com isopropanol antes de secos por centrifugação imediata a 1000 rpm durante 8 a 10 minutos. Amostras e controlos foram combinados de forma a hibridar a mesma quantidade de DNA marcado. Os tampões de hibridação e lavagem da Agilent (Oligo aCGH/CHIP-onChIP Hybridization/Wash Buffer kit) foram utilizados, seguindo as instruções presentes em *Agilent Oligonucleotide Array-Based CGH for Genomic DNA Analysis Protocol (version 4.0)*. Contudo, a duração das hibridações foi reduzida de 40 para 20 horas. Todas as amostras foram hibridadas em duplicado correspondendo às réplicas técnicas: numa hibridação a amostra foi marcada com Cy5 e o controlo com Cy3, noutra hibridação a amostra foi marcada com Cy3 e o controlo com Cy5.

#### 4. Aquisição e análise de dados

Os slides lavados e secos foram lidos num scanner de microarrays Agilent G2565AA. A intensidade de fluorescência dos sinais detectados foi quantificada e extraída utilizando o software QuantArray v3.0 (Packard BioScience). Os dados foram filtrados do seguinte

modo. Foram assinalados manualmente os “maus *spots*”<sup>1</sup>, eliminando-se assim *spots* cuja impressão tenha sido incorrecta ou nos quais tenha sido detectada alguma imperfeição ou impureza que afecte a fluorescência de um dos canais. Genes cuja razão de intensidade sinal/intensidade do sinal de fundo fosse inferior a  $1,5 \pm 0,5$  foram também eliminados. O sinal de fundo foi subtraído localmente a cada spot e a média do sinal das réplicas de cada gene foi também calculada. A normalização dos dados foi feita com o software BRB-Arrays Tools (<http://linus.nci.nih.gov/BRB-ArrayTools.html>). A mediana de cada array, assim como a transformação logarítmica (base 2) das razões Cy5/Cy3, foi calculada para cada sonda.

O *cluster* hierárquico (agrupamento hierárquico) entre estirpes foi efectuado com os dados filtrados e normalizados no BRB-Array Tools. Este agrupamento das estirpes foi efectuado com correlação centrada de Pearson e *Average linkage*.

O software CGH-Miner (Wang *et al.*, 2005) foi utilizado seguindo as instruções presentes no manual do utilizador para gerar os gráficos consenso. Tal como descrito por Dunn e colaboradores (2005), para uma menor janela de análise dos dados optou-se pelo parâmetro *BAC-arrays*. Para a visualização dos perfis cariotípicos, utilizou-se o software Java TreeView (Eisen *et al.*, 1998).

A identificação de genes significativamente diferentes nos subgrupos de estirpes, definidos *a-priori* com base no agrupamento hierárquico, foi feita com a técnica SAM (Tusher *et al.*, 2001). Para cada subgrupo foi definido um valor de  $\delta$  que correspondesse a uma FDR (*False Discovery Rate*) = 0.

O *bi-clustering* foi efectuado com o algoritmo xMotif (Murali & Kasif, 2003) através da plataforma bicAT (Barkow *et al.*, 2006).

---

<sup>1</sup> *Spots*: Elementos da matriz do *microarray* no qual foi hibridada uma amostra de DNA.





## **RESULTADOS**

## 1. Arrays de gDNA

Os *arrays* de gDNA obtidos nas hibridações de aCGH tiveram aspecto semelhante ao representado no Apêndice 1. A maioria dos *spots* apresentaram cor amarela, alguns *spots* apresentaram cor vermelha ou verde. Diferentes fluoróforos tiveram diferentes eficácias de marcação do gDNA deslocando a intensidade global de cada *array* num determinado sentido, existindo uma relação não linear entre as duas cores utilizadas (Cy3 e Cy5), não só nas experiências deste estudo mas em praticamente todas as experiências de *microarrays* com duas cores. Para controlar e compensar este fenómeno, experiências de troca de fluoróforos entre amostras de DNA foram efectuadas em simultâneo. A média das experiências de troca de fluoróforos elimina tais artefactos sendo representativa das alterações reais observadas em cada amostra.

Os *spots* de clones genómicos cuja impressão tenha sido incorrecta (impurezas, arrastamentos, bolhas de ar, etc.) foram assinalados inicialmente e eliminados da análise dos resultados (Figura 8A). Também os “maus *spots*” pela presença de impurezas durante a hibridação e lavagem foram eliminados na filtração manual (Figura 8B).

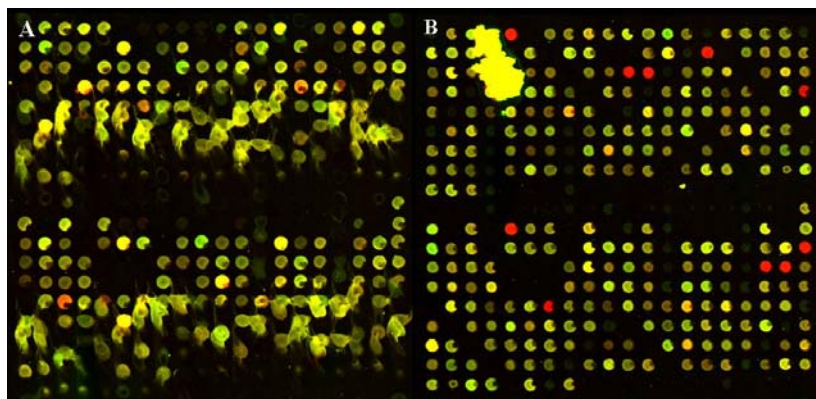


Figura 8. Exemplos de "maus *spots*" obtidos pontualmente nos *arrays* de aCGH.

A intensidade de cada clone genómico hibridado é o somatório da intensidade do ruído e da amostra de DNA marcado e, portanto, para obter apenas o valor correspondente ao DNA hibridado eliminou-se o ruído de fundo. Tal, foi realizado subtraindo à intensidade de cada *spot* a intensidade do ruído adjacente (correspondente à área extra de um quadrado com um círculo inscrito). Este método local é vantajoso pois tem em consideração as

variações no ruído por toda a matriz do *array*. Também os *spots* cuja razão entre a intensidade de sinal e a intensidade de ruído de fundo era inferior  $1,5 \pm 0,5$  foram eliminados. Esta filtração, embora bastante permissiva, elimina genes cujo sinal era apenas ruído na amostra experimental uma vez que o sinal de fundo e o sinal da amostra pouco diferiam. Calcularam-se finalmente médias dos valores das réplicas de cada sonda num *array* para obter um único valor representativo do comportamento global de cada gene. Uma vez que cada amostra foi hibridada em duas réplicas técnicas e que, em cada uma destas, cada sonda estava impressa em duplicado o valor final de cada ORF é a média de quatro valores experimentais.

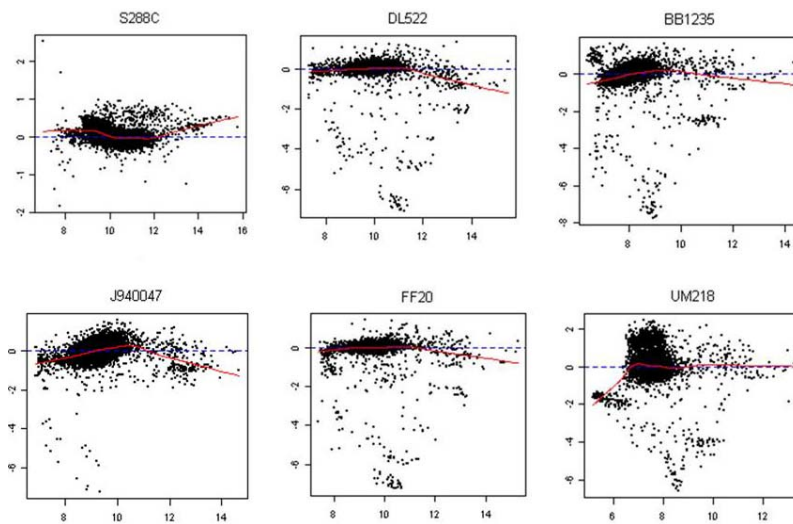
A fim de se poderem comparar experiências/estirpes de forma independente, a normalização dos resultados foi essencial. Uma vez que não se esperavam alterações de todo o genoma de uma estirpe, mas apenas pequenas diferenças do número de cópias de alguns fragmentos cromossómicos ou até de alguns cromossomas (aneuploidias ou poliploidias), todos os dados foram centrados pela mediana da intensidade de sinal de cada *array*. A mediana foi escolhida em detrimento da média pois é menos sensível ao efeito de *outliers*<sup>2</sup>. Uma vez que o resultado da normalização de todos os *arrays* pela mediana do sinal de cada um, e da normalização pela mediana e pela escala de intensidades dos *arrays* não apresentou grande diferença, apenas o primeiro tipo de normalização foi efectuado para evitar manipulação excessiva e desnecessária dos dados. No Apêndice 2 pode ver-se o efeito da normalização dos *arrays*.

A transformação logarítmica das razões das duas cores efectuou-se para tornar a interpretação biológica dos resultados mais evidente. Por exemplo, um dado gene numa amostra-**A** tem valor absoluto de 10,000 relativamente ao mesmo gene na referência-**A** 1,000. O mesmo gene numa amostra-**B** (a mesma amostra numa outra hibridação) tem valor absoluto 1,000 relativamente à referência-**B**, 100. Em valores absolutos, dir-se-ia que este gene está *aumentado* na amostra-**A** ( $10,000 - 1,000 = 9,000$ ) do que na amostra-**B** ( $1,000 - 100 = 900$ ). No entanto, a diferença em termos biológicos é a mesma: um aumento de 10 vezes. A transformação logarítmica simplifica estes raciocínios. Ou seja,  $\log_{10}(10000) - \log_{10}(1000) = \log_{10}(1000) - \log_{10}(100)$ . Assim, a aplicação de escala logarítmica torna comparáveis as diferenças de valores em diferentes intensidades.

---

<sup>2</sup> *Outlier*: Ponto correspondente a sonda cuja hibridação tem sinal distinto da maioria das restantes sondas.

Na referência, para a qual foi desenhada a colecção de sondas, a maioria dos sinais encontrava-se numa nuvem centrada em  $M = 0$ ;  $[M = \log_2 (\text{vermelho/verde}) = 0]$ , pois as intensidades do sinal vermelho e verde são iguais, daí a cor amarela. Para as restantes estirpes, embora grande parte dos sinais se concentrassem em  $M = 0$ , havia alguns *spots* localizados fora destas nuvens. Esta distribuição significou que, tal como esperado, embora grande parte das ORFs destas leveduras fossem conservadas relativamente ao controlo (estirpe S288C; nuvem central), algumas apresentavam alterações no número de cópias. Estas diferenças puderam visualizar-se nos gráficos MA ( $M = f(A)$ ,  $A = [\log_2 (\text{Cy5}) + \log_2 (\text{Cy3})] / 2$ ) da Figura 9 e nos *Scatter plots* do Apêndice 3.



**Figura 9. Gráficos MA para uma estirpe autóctone *S. cerevisiae* representativa de nichos ecológicos diferentes.**

Os pontos cujo  $M < 0$  representam ORFs com baixa homologia ou menor número de cópias e os pontos cujo  $M > 0$  representam ORFs com maior número de cópias.  $A = [\log_2 (\text{Cy5}) + \log_2 (\text{Cy3})] / 2$ ;  $M = \log_2 (\text{Cy5/Cy3})$ .

## 2. Estirpes de *S. cerevisiae*.

As quinze estirpes autóctones de *S. cerevisiae* utilizadas neste estudo provêm de nichos ecológicos diferentes. Quatro são de origem comercial, quatro foram recolhidas em lagares da Bairrada durante a fermentação de mostos, quatro provêm de vinhas, duas da Bairrada e duas da Região Demarcada do Vinho Verde e três foram isolados clínicos provenientes de indivíduos imuno-deprimidos. As estirpes de origem comercial, Davis Lalvin 522, IOC 18-2007, AEB Fermol Rouge e ICV D254, são estirpes industriais utilizadas na fermentação de diferentes tipos de vinhos.

**Tabela 5. Características dos vinhos e estilos de fermentação mais adequados às estirpes de *S. cerevisiae* industriais analisadas neste estudo**

	<b>Davis Lalvin 522</b>	<b>IOC 18 – 2007</b>	<b>AEB Fermol Rouge</b>	<b>ICV D254</b>
	<i>S. cerevisiae</i> estirpe M-1107	<i>S. cerevisiae</i> Killer	<i>S. cerevisiae</i>	<i>S. cerevisiae</i> , estirpe D254
Estilo usado	Tintos e brancos com corpo completo; frutos de Outono	Tintos com corpo completo e pesado e broncos robustos; Fermentação em garrafa (vinhos espumantes de qualidade)	Tintos e Roses; vinificações e fermentações de uvas que iniciaram já maceração carbónica	Tintos Roses Mediterrânicos, Syrah (especialmente fermentações em barris); Brancos e Roses fermentados em barris, uvas muito maduras
Outras características	Boa extracção de cor, intensa e complexa; cuidado com temperaturas elevadas	Produção glicerol 6 g/l	Rapidamente se sobrepõe a flora autóctone; curto período de latência; inicia fermentação rapidamente	Boa sinergia com vinhos brancos produzidos com ICV D80 e ICV D21

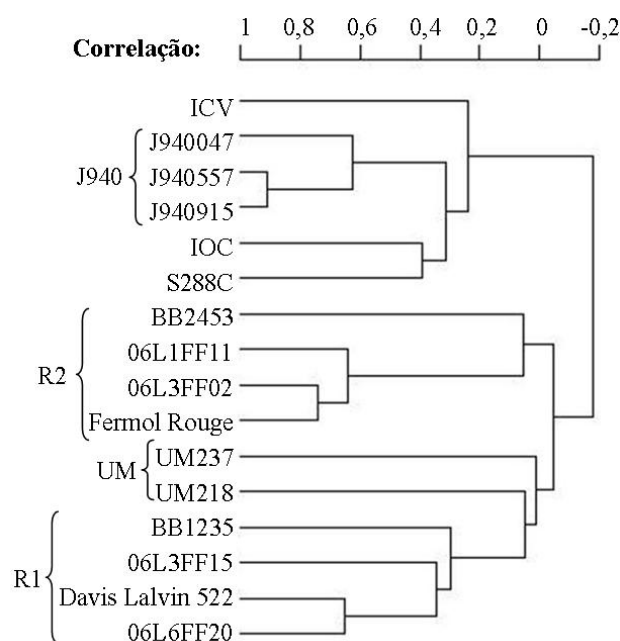
As estirpes isoladas de lagares da Bairrada foram recolhidas em diferentes locais (L1, L3 e L6) durante o ano de 2006 (06). Foram seleccionadas estirpes de quatro diferentes processos de fermentação fina (FF02, FF11, FF15, FF20). As estirpes recolhidas nas vinhas têm duas diferentes origens geográficas, duas são da Região Demarcada da Bairrada (**Baga Bairrada**) e duas são da Região Demarcada do Vinho Verde. As estirpes da Região Demarcada da Bairrada foram recolhidas em diferentes vinhas (1 e 2) e escolhidos 2 clones diferentes (235 e 453). As duas restantes foram cedidas pelo Laboratório de Biotecnologia de Leveduras do Departamento de Biologia da Universidade do Minho, UM218 e UM237. Finalmente, as três estirpes que foram isoladas de indivíduos imuno-deprimidos (J940047 e J940557 e J940915), no condado de Kent em Inglaterra.

Embora de diferentes origens, todas as estirpes autóctones da Região Demarcada da Bairrada, as quatro estirpes de origem comercial/industrial e as três estirpes isoladas de indivíduos imuno-deprimidos são *S. cerevisiae*, tendo sido confirmado pelo padrão de restrição do locus *MET2* (**METH**ionine requiring) com EcoRI e por sequenciação do rDNA 18S e ITS efectuado pela Unidade de Genómica do Biocant (Anexo 1). As estirpes dos lagares de fermentação provêm da Região Demarcada da Bairrada. Todas foram recolhidas durante o ano de 2006 (06) mas em diferentes lagares (L1, L3 e L6). Foram seleccionadas estirpes de quatro diferentes processos de fermentação fina (FF02, FF11, FF15, FF20).

### 3. Agrupamento hierárquico das estirpes de *S. cerevisiae*

Uma vez que a correlação de Pearson depende apenas do padrão de todos os elementos de uma variável e não da intensidade do sinal dos elementos, foi o método escolhido para calcular a distância entre elementos. De referir que em agrupamentos hierárquicos, a posição de determinado elemento (estirpe) num grupo não traduz qualquer significado no espaço, apenas maior semelhança com os elementos desse mesmo grupo e maior dissimelhança com os elementos de outro grupo. A distância entre ramos/grupos foi calculada através do método *Average linkage*, que calcula a média de todas as distâncias entre os elementos de cada grupo (Draghici, 2003) (Figura 10).

Quatro hibridações de S288C simultaneamente como amostra e referência (*self-self*) serviram para gerar uma linha de base para a análise. Assim, por comparação de cada estirpe com a mesma referência, a comparação entre as diferentes estirpes pôde ser efectuada.



**Figura 10. Agrupamento hierárquico das 15 estirpes autóctones de *S. cerevisiae* analisadas juntamente com a estirpe de referência *S. cerevisiae* S288C.**

Agrupamento hierárquico de estirpes foi feito usando software BRB-Array Tools, com base nos valores de M filtrados e normalizados para todas as amostras utilizando a correlação de Pearson e o método average linkage. As chavetas e os respectivos nomes referem-se a subgrupos de estirpes autóctones considerados em análises subsequentes.

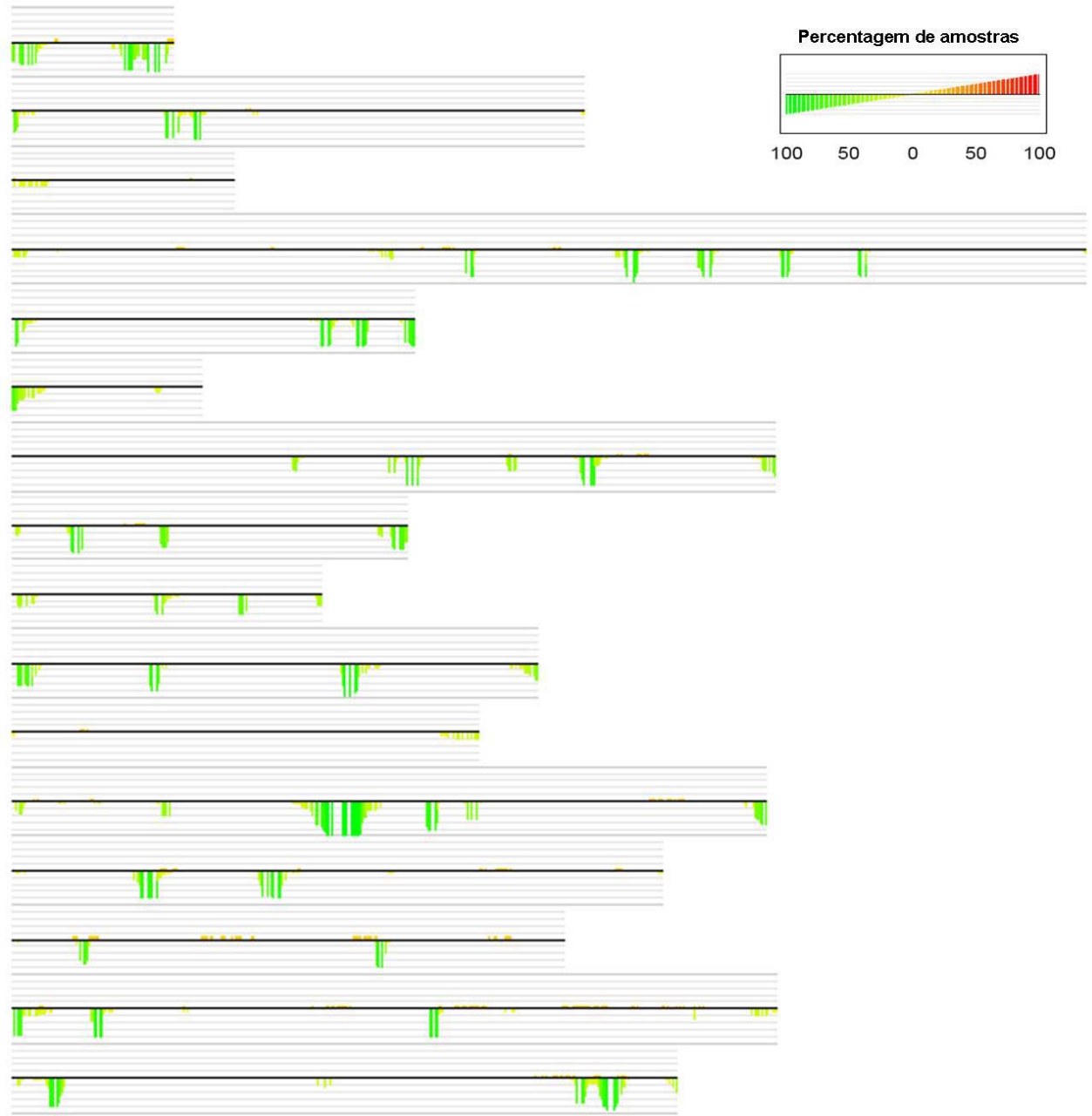
Embora com origens diferentes, as estirpes isoladas das vinhas e lagares não se organizaram nem separaram de forma relacionada com a sua proveniência. As estirpes clínicas formaram, em conjunto com IOC e ICV um grupo hierarquicamente afastado das restantes. A maior parte das estirpes de origem vitivinícola, tanto da Região Demarcada do Vinho Verde como da Região Demarcada da Bairrada organizaram-se num outro ramo bem distinto do formado pelas estirpes anteriormente referidas.

#### **4. Cariótipos e consenso das alterações cromossómicas**

Os cariótipos das estirpes gerados pelo software CGH-Miner representam graficamente as regiões com alterações cromossómicas relativamente a *S. cerevisiae* S288C. Os resultados obtidos para os cariótipos e consensos de alterações cromossómicas não reflectiram qualquer alteração isolada de um único gene, pois este software construiu todas as representações gráficas com base numa janela de 3 genes, da qual considerou a média como valor do respectivo segmento cromossómico. Na Figura 11 encontra-se representado o gráfico consenso das alterações cromossómicas de todas as estirpes autóctones *S. cerevisiae* estudadas. Na posição referente a cada ORF há uma barra que corresponde à percentagem de amostras cuja alteração do número de cópias foi significativa relativamente a S288C. As barras vermelhas representam regiões amplificadas e as barras verdes regiões com menor número de cópias. Cada linha horizontal cinzenta representa 20% das estirpes analisadas.

As regiões de maior variabilidade genética foram definidas como ORFs que variaram em 75% das estirpes autóctones estudadas Dunn *et al.* (2005). Ou seja, apenas as ORFs que apresentaram alterações cromossómicas em pelo menos 11 das 15 estirpes *S. cerevisiae* foram consideradas como fazendo parte do consenso das ORFs de maior variabilidade genética. Nos perfis cariotípicos de cada estirpe (Apêndice 3), cada ORF foi colocada na respectiva localização cromossómica. A altura de cada barra verde ou vermelha é proporcional à razão  $\log_2$  (Cy5/Cy3). Se esta razão for inferior a 1 é desenhada abaixo da linha de base e assinalada a verde; assim, cada barra verde representa uma deleção, menor número de cópias relativamente à referência ou baixa homologia. Se a razão for superior a 1, a barra é colocada acima da linha de base e colorida a vermelho, o que caracteriza uma amplificação ou maior número de cópias relativamente a *S. cerevisiae* S288C. Qualquer

alteração cromossômica numa estirpe autóctone foi detectada pelo desvio da razão esperada 1:1 (amostra:referência) uma vez que se assumiu que as estirpes em estudo não divergiram mais do que 25% da estirpe laboratorial (Masneuf *et al.*, 1998).



**Figura 11. Gráfico consenso das alterações cromossômicas verificadas nas estirpes autóctones de *S. cerevisiae* analisadas.**

O programa CGH-Miner (Wang *et al.*, 2005) gera gráficos consenso das alterações cromossômicas das estirpes autóctones de *S. cerevisiae* analisadas relativamente à estirpe de referência, *S. cerevisiae* S288C. Os 16 cromossomos estão dispostos por ordem numérica do I ao XVI. Na posição referente a cada ORF encontra-se uma barra que corresponde à percentagem de amostras cuja alteração do número de cópias foi significativa relativamente a S288C. As barras vermelhas representam regiões amplificadas e as barras verdes regiões com menor número de cópias. Cada linha cinzenta representa 20% das estirpes analisadas.



Das 6381 ORFs analisadas verificaram-se alterações cromossómicas em 1276 ORFs. Ou seja, tal como se esperava, apenas 20% das ORFs de todo o genoma das leveduras apresentaram alterações. Destas, apenas foram consideradas como altamente variáveis 188 ORFs alteradas em 75% das estirpes de *S. cerevisiae*. Excluíram-se, ainda, ORFs sem anotações disponíveis, definidas como ORFs hipotéticas ou dúbias, e a assinatura de maior variabilidade genética resumiu-se a 147 ORFs. Na Figura 12, está representada a distribuição destas ORFs por processo biológico. A maioria destas está relacionada com processos metabólicos. Seguem-se a organização e biogénese celular, a regulação dos referidos processos biológicos e a transposição. Estes genes, alterados em pelo menos 11 das 15 estirpes autóctones de *S. cerevisiae*, distribuem-se nas categorias de processos biológicos referidos, sendo as três primeiras categorias redundantes, uma vez que as ORFs de cada uma destas categorias são praticamente as mesmas.



**Figura 12. Distribuição das ORFs com alteração do número de cópias em pelo menos 11 das 15 estirpes autóctones de *S. cerevisiae* por processo biológico.**

Representação gerada pela ferramenta FatiGO com base na lista das ORFs com diferente número de cópias em pelo menos 11 das 15 estirpes autóctones de *S. cerevisiae*. Os valores das percentagens indicam o número relativo de ORFs em cada categoria (Al-Shahrour et al., 2004; -, 2005; -, 2006).

As três categorias de processos metabólicos estavam maioritariamente associadas a proteínas de ligação de nucleótidos ou ácidos nucleicos. Os genes destas categorias detectados em menor número de cópias referem-se a factores envolvidos no *splicing* de tRNA, helicases subteloméricas, factores de maturação de rRNA (25S ou 5.8S), proteínas que asseguram a correcta segregação cromossómica, proteínas que reconhecem e se ligam a DNA danificado nos mecanismos de remoção de nucleótidos, factores de iniciação da elongação, reguladores da transcrição (associado a *PDR1*, *Pleotropic Drug Resistance*),

componentes de transcrição de rDNA. A ligação a outras proteínas, também bastante afectada pelo decrescente número de cópias nas estirpes, reporta-se a proteínas de ligação à proteína 14-3-3, à actina ou a transportadores de membrana. As categorias de genes onde foram encontradas muitas ORFs com diferente número de cópias representam famílias de genes ou apresentam homologia com outros elementos do genoma.

Relativamente aos restantes alelos mais variáveis, destacaram-se os que participam na organização e biogénese celular, na regulação de processos biológicos, os alelos associados à localização celular, maioritariamente constituído por transportadores e genes de resposta ao stress.

Na tabela 6 estão alguns exemplos de genes destas categorias, respectivas funções e possíveis alterações fenotípicas resultantes do diferente número de cópias.

**Tabela 6. Exemplos de genes muito variáveis nas estirpes *S. cerevisiae* autóctones deste estudo.**

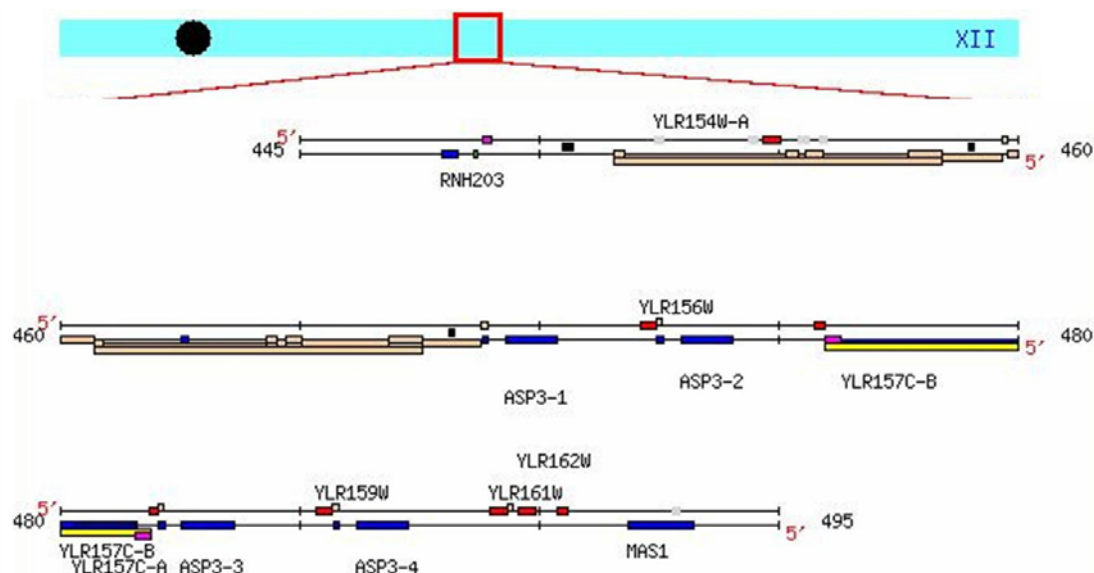
	<b>Crom.</b>	<b>Função descrita</b>	<b>Possível alteração fenotípica</b>
<i>SWH1</i>	I	Proteína de ligação a oxisteróis (metabolitos biossintéticos de esteróis) na formação de vesículas do Complexo Golgi	Embora não essencial à célula, a outra cópia de <i>SWH</i> permaneceu no genoma não se esperando grandes alterações; possível alteração na resistência à nistatina
<i>SEN34</i>	I	( <i>Splicing ENdonuclease</i> ) Uma das quatro subunidades da endonuclease envolvida no splicing de tRNAs	Poderá estar bloqueada a clivagem a 5' do local onde o intrão seria retirado
<i>PDR3, 11 e 12</i>	II, IX, XVI	( <i>Pleiotropic Drug Resistance</i> ) Associados à resistência múltipla a drogas, regulação da expressão de transportadores da casset de ligação ao ATP	Outros genes da família permaneceram no genoma, possível diminuição da resistência a drogas
<i>ESF1</i>	IV	( <i>Eighteen s rRNA Factor</i> ) Proteína nuclear envolvida no processamento pre-rRNA	Diminuição possivelmente acentuada de níveis de 18S rRNA
<i>GCD6</i>	IV	( <i>General Control Derepressed</i> ) Em conjunto com eIF-2B controla tradução de <i>GCN4</i> ( <i>General Control Nonrepressible</i> )	Mais nenhum factor envolvido na tradução de <i>GCN4</i> foi afectado, poderá não haver alterações na resposta ao défice de azoto
<i>GDI</i>	V	( <i>GDP Dissociation Inhibitor</i> ) Componente da regulação de GDPs (proteínas da família das GTPases) no tráfego de vesículas entre organelos	Bloqueio parcial do transporte nas vias secretoras e biossíntese de vacuolos
<i>YRF1-2</i>	V	Helicase subtelomérica (Y'-HELicase Protein 1, Help1p)	Aumento da recombinação homologa DNA, diminuição da longevidade das células
<i>ORM1</i>	VII	Resistência a agentes que induzem resposta de proteínas sem correcta estrutura 3D	Acumulação de proteínas com estrutura 3D incorrecta
<i>TIF4631</i>	VII	( <i>Translation Initiation Factor</i> ) Homólogo de eIF4G, factor de iniciação de tradução, associa-se simultaneamente a eIF4E e Pab1p	Região semelhante a Pab1p parece ser essencial ao crescimento celular; eIF4E permaneceu no genoma, não se esperariam danos na tradução.
<i>STM1</i>	XII	( <i>Supressor os tom1</i> ) Em conjunto com Cdc13p mantém estrutura de regiões subteloméricas	Alterações na estrutura de telómeros e Y'
<i>PUS5</i>	XII	( <i>PseudoUridine Synthase</i> ) Envolvido na metilação de 21S rRNA mitocondrial	4 restantes cópias do gene <i>PUS</i> inalteradas, não se esperam grandes efeitos
<i>CSM3</i>	XIII	( <i>Chromosome Segregation in Meiosis</i> ) Necessária à correcta segregação de cromossomas	Possíveis alterações na segregação cromossómica
<i>PRM6</i>	XIII	( <i>Pheromone-Regulated Membrane protein</i> ) Regulado por Ste12p durante reprodução sexuada	Menor facilidade de crescimento
<i>RRN11</i>	XIII	( <i>Regulation of RNA polymerase I</i> ) Componente de complexo essencial à transcrição de rDNA,	DNA ribossomal poderá ser menos abundante nestas células
<i>YMD8</i>	XIII	Semelhante a transportador essencial à glicosilação de proteínas e lípidos no Complexo de Golgi, Vrg4p	<i>VRG4</i> permaneceu inalterado, não se esperam grandes alterações na glicosilação

<i>POR1</i>	XIV	( <i>PORin</i> ) Porina mitocondrial	Maior permeabilidade iónica, desestabilização osmótica da mitocôndria
<i>ITR2</i>	XV	( <i>myo-Inositol Transporter</i> ) Transportador de inositol para as células, efectivo com várias cópias	<i>ITR1</i> permanece no genoma não permitindo desequilíbrio de importação de inositol para a célula
<i>NDJ1</i>	XV	( <i>NonDisJunction</i> ) Proteína telomérica específica de meiose: emparelhamento de cromossomas homólogos, correcto <i>crossing-over</i> e segregação de pequenos cromossomas	Erros na segregação cromossómica
<i>TPT1</i>	XV	( <i>tRNA 2'-phosphotransferase</i> ) tRNA 2'-fosfotransferase envolvida no <i>splicing</i> de tRNA	<i>Splicing</i> de tRNAs poderá ser menor quando comparado com <i>S. cerevisiae</i> S288C.
<i>WSC3</i>	XV	( <i>cell Wall integrity and Stress response Component</i> ) Sensor na sinalização activa de PKC1-MPK1; Associado à integridade da membrana celular	Síntese de 1,3- $\beta$ -glucano poderá ficar comprometida
<i>MEP3</i>	XVI	Permease de amónia de baixa afinidade (activada pela ausência de Azoto), permite efluxo de amónia para espaço extracelular como fonte alternativa de Azoto	Gln3p e Nil1p continuarão a regular <i>MEP1</i> e <i>MEP2</i> não se verificando grandes alterações

Os elementos *Ty* evidenciaram-se por representarem aproximadamente 18% destas ORFs mais variáveis. Tal como mencionado anteriormente, são sequências de DNA que se movimentam ao longo dos genomas recorrendo a intermediários de RNA (Roeder & Fink, 1982). Em todas as estirpes autóctones de *S. cerevisiae* de origem vitivinícola estes fragmentos de elevada mobilidade foram detectados em menor número de cópias do que em *S. cerevisiae* S288C. Esta diminuição do número de elementos *Ty* é concordante com o descrito na literatura (Hauser *et al.*, 2001; Dunn *et al.*, 2005), segundo a qual este facto estará relacionado com uma possível pressão negativa sobre a acumulação de transposões durante a fermentação alcoólica que causa a perda destes elementos (Jordan & McDonald, 1999). Em conjunto, os telómeros e os elementos *Ty*, fornecem redundância aos cromossomas impedindo a perda de informação fundamental codificada em ORFs adjacentes a estas regiões, aquando da recombinação ectópica ou mesmo segregação cromossómica (Kupiec & Petes, 1988; Moore & Harber, 1996). Assim, a eliminação destes elementos previne a eliminação de fragmentos cromossómicos fundamentais para a sobrevivência da célula.

À medida que as percentagens de genes que estão diminuídos no seu número de cópias se tornam mais pequenas, o número de genes por categoria diminui. Esta reduzida perda de cópias nas referidas categorias, significa que as respectivas funções biológicas são de tal modo fundamentais para a célula que a perda de genes nelas envolvidos, não confere vantagens evolutivas, mas sim desvantagens. Assim, genes envolvidos em processos essenciais como a autofagia, crescimento filamentoso, segregação cromossómica ou desenvolvimento de estruturas anatómicas são exemplos de processos biológicos pouco variáveis no genoma destas leveduras selvagens.

Uma análise mais detalhada das ORFs pertencentes ao consenso das estirpes de *S. cerevisiae* deste estudo permitiu verificar que das referidas ORFs, apenas 14 são comuns a todas as estirpes autóctones. Estas 14 ORFs estão localizadas no cromossoma XII, como representado na figura 14. No centro deste segmento encontram-se um elemento *TyA* e um elemento *TyB* ladeados de 2 das 4 cópias da Asparaginase II (*ASP3*, *ASParaginase*). As restantes 2 cópias desta asparaginase estão alternadas com YLR156W e YLR159W. Acrescentam-se a esta pequena lista *ACS2* (*Acetyl CoA Synthetase*), *RNH203* (*RNase H*) *MAS1* (*Mitochondrial Assembly*).



**Figura 13. Mapa da localização relativa de algumas ORFs cujo número de cópias é diferente relativamente à referência *S. cerevisiae* S288C em todas as estirpes autóctones.** Cromossoma XII, 445000-495000. Assinaladas a azul estão *RNH203*, as quatro cópias de *ASP3*, *MAS1*. Os elementos Ty estão a azul-escuro e amarelo (Adaptado de SGD).

O inferior número de cópias das quatro asparaginases II era já esperado para todas as estirpes (Edwards-Ingram *et al.*, 2004; Hauser *et al.*, 2001; Lashkari *et al.*, 1997; Winzeler *et al.*, 2003) pois esta deleção tem vindo a ser verificada em praticamente todas as estirpes *S. cerevisiae* que não S288C. Uma vez que todas as cópias desta asparaginase estão em muito menor número de cópias do que na referência S288C poderia observar-se ausência desta glicoproteína nas células destas leveduras. Sabe-se que esta enzima da parede celular, envolvida no catabolismo da D- e L-asparagina em aspartato e amónia (Dunlop *et al.*, 1978), possui uma isoforma citosólica, a asparaginase I. Verificou-se que esta última isoforma permaneceu no genoma das leveduras autóctones estudadas. Assim, se um possível défice de azoto induzisse a actividade das asparaginases (Kim *et al.*, 1988), amónia continuaria a ser gerada com fonte substituta de azoto. Pelo facto de ao longo dos anos se ter verificado esta constância quanto ao desaparecimento das asparaginases II, esta alteração cromossómica poderá estar já bem fixa no processo evolutivo não só de estirpes vínicas como em outras estirpes *S. cerevisiae*.

A isoforma ACS2 da acetil-coA sintetase é a fonte de acetil-coA na acetilação de histonas. Particularmente expressa sob condições de anaerobiose, é necessária para o crescimento das células na presença de glucose. Uma vez que, embora em menor número de cópias, esta não é a única via de produção de acetil-coA, o complexo piruvato desidrogenase mitocondrial (PDH) e a  $\beta$ -oxidação de ácidos gordos assegurarão a sua produção. Esta isoforma nuclear, ACS2, tem também como função a regulação da expressão genética (Hall *et al.*, 2004). Visto que a acetil-coA deixou de ser fornecida no núcleo à acetilação de histonas, espera-se que nestas estirpes autóctones de *S. cerevisiae* a expressão genética esteja diminuída, a acetilação de histonas parece ser necessária à transcrição global (Takahashi *et al.*, 2006).

*RNH203* codifica uma subunidade da ribonuclease H2 que fundamental ao seu funcionamento uma vez que parece ser necessária para a conformação activa da RNaseII (Jeong *et al.*, 2004). Uma vez que este gene apresentou um pequeno decréscimo do número de cópias em todas as estirpes autóctones de *S. cerevisiae*, este facto poderia indicar um aumento da taxa de mutações nestas estirpes (Jeong *et al.*, 2004). No entanto, uma vez que outra RNaseH (H1) permanece inalterada, esta impedirá a acumulação de mutações.

Acrescenta-se a este balanço o papel da Helicase Sgs1p na resolução dos possíveis erros causados pela inoperatividade da RNaseH2 (Ooi *et al.*, 2003).

A importação de proteínas para a mitocôndria depende da clivagem de uma sequência sinalizadora por uma protease mitocondrial (McAda & Douglas, 1982). Esta protease é constituída por duas subunidades (cujos genes são *MAS1* e *MAS2*) habitualmente nas células em concentrações equimolares. O gene para a subunidade menor desta protease (*MAS1*) teve um decréscimo do número de cópias em todas as estirpes autóctones de *S. cerevisiae* relativamente à referência, enquanto o gene da subunidade maior (*MAS2*) não sofreu qualquer alteração. UM218 é neste último caso uma excepção pois foram detectadas 3 cópias a mais relativamente a S288C. Embora, a activação da protease esteja comprometida (Witte *et al.*, 1988) não será de espera um total bloqueio da importação de proteínas para a mitocôndria (Yang *et al.*, 1988).

Na tabela 7 estão contabilizadas as ORFs com diferente número de cópias em cada estirpe selvagem relativamente a *S. cerevisiae*.

**Tabela 7. Número total de ORFs detectadas em diferente número de cópias em cada estirpe *S. cerevisiae* autóctone relativamente a S288C.**

<b>Estirpe</b>	<b>Número de ORFs com diferente número de cópias</b>
Davis Lalvin 522	387
IOC 18 – 2007	346
AEB Fermol Rouge	352
ICV D254	337
06L3FF02	378
06L1FF11	489
06L3FF15	514
06L6FF20	425
BB1235	451
BB2453	347
UM218	318 (+ 336*)
UM237	249
J940 047	111
J940 557	170
J940 915	136

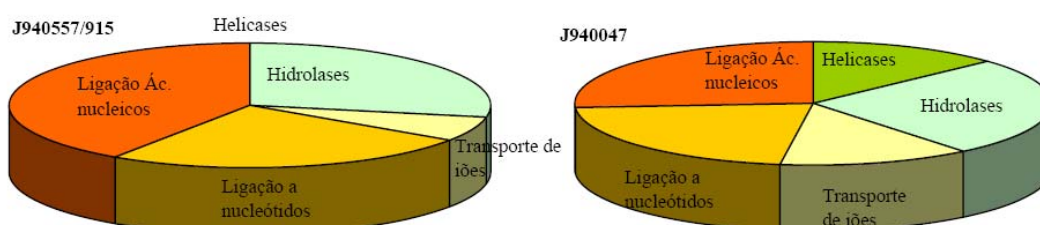
Em geral, todas as estirpes selvagens têm um número de ORFs inferior à estirpe controlo (S288C). UM218 foi a única estirpe em que as alterações cromossómicas verificadas quase se igualaram nos dois sentidos, ou seja, 318 ORFs foram detectadas em menor número de



cópias relativamente a S288C, e 336 (\*) ORFs foram encontradas em maior número de cópias relativamente a *S. cerevisiae* S288C.

## 5. Alterações cromossómicas de subgrupos hierárquicos de estirpes selvagens

As estirpes clínicas formaram um grupo com 60% de correlação entre si sendo que J940557 e J940915 são mais semelhantes (correlação ~ 90%). J940047 foi, das três estirpes, a que apresentou menores alterações relativamente a S288C, tendo por isso sido colocada separadamente das outras duas.



**Figura 14. Representação gráfica das ORFs onde se verificou alteração no número de cópias em J940557 e J940915 como subgrupo, relativamente a J940047. Distribuição por categoria de função molecular.**

As categorias de funções moleculares foram definidas pelo FatiGO, assim como as respectivas percentagens relativas (Al-Shahrour et al., 2004; -, 2005; -, 2006).

Enquanto o subgrupo de J940557/J940915 não perdeu genes codificantes de helicases, J940047 perdeu três cópias do gene *YRF1* (2, 4 e 5) que codifica uma helicase associada aos telómeros, e três ORFs que hipoteticamente estão associadas a esta actividade. As restantes categorias de funções moleculares foram idênticas nos dois grupos de estirpes variando a percentagem relativa de ORFs e os genes que, em particular, formaram as ditas categorias. Acrescente-se ainda que no subgrupo J940557/J940915 foram detectados alguns genes em inferior número de cópias relativamente a S288C que não foram detectados em J940047 (Tabela 8).

**Tabela 8. Exemplo de genes apenas em inferior número de cópias em J940557/J940915 comparativamente com J940047.**

<i>HXT15</i> <i>CUP1</i>	<i>HeXose Transporter</i>	Transportador de hexoses da membrana
		Proteína de ligação ao cobre associada aos mecanismos de resistência a elevadas concentrações deste metal e cádmio
<i>SIR1</i>	<i>Silent Information Regulator</i>	Envolvido no silenciamento dos <i>locus HML</i> e <i>HMR</i> da reprodução sexuada
<i>FLO10</i> <i>NFT1</i>	<i>FLOcculation</i> <i>New Full-length MRP-type Transporter</i>	Pertencente à família de floculinas <i>FLO</i> Codifica um transportador associado à família de proteínas de resistências múltiplas a drogas
<i>MPH2</i>	<i>Maltose Permesase Homolog</i>	Codifica um transportador de maltase

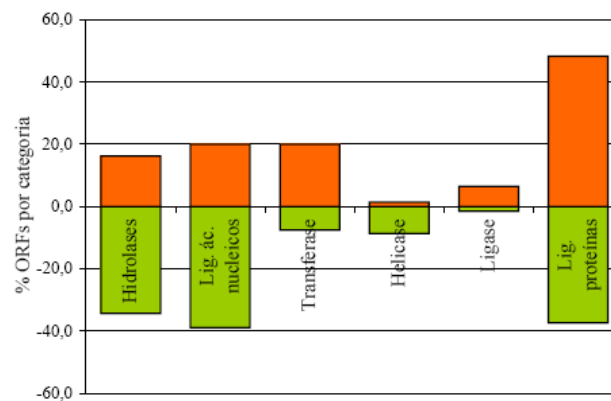
Uma característica fenotípica das estirpes isoladas de indivíduos imuno-deprimidos é a sua capacidade de secreção de proteínas extracelulares. *GDII* (*GDP Dissociation Inhibitor 1*), tal como já referido é um componente da regulação de GDPs, proteínas da família das GTPases importantes no tráfego de vesículas entre organelos. Uma vez que as estirpes clínicas não perderam nenhuma das cópias destes genes, ao contrário do que foi observado para as estirpes de origem vitivinícola, as primeiras terão maior capacidade secretora. O gene *PDII* (*Protein Disulfide Isomerase*), que codifica a bissulfato isomerase do lúmen do retículo endoplasmático é responsável pela formação de ligações bissulfato em proteínas da superfície celular ou secretadas. Apenas a estirpe J940915 perdeu este gene, o que significa que as restantes duas estirpes deste subgrupo estarão mais adaptadas aos ambientes que habitualmente colonizam, tendo maior facilidade no processo de secreção. Este gene foi o único que J940915 perdeu e J940557 manteve. Todas as outras alterações cromossómicas de J940915 são partilhadas com J940557.

As duas estirpes comerciais IOC e ICV foram as de origem vitivinícola mais semelhantes a S288C, pertencendo ao mesmo ramo hierárquico que S288C. Esta proximidade não está relacionada com o número de diferenças encontradas relativamente à estirpe referência, uma vez que UM237 possui menos alterações do que as estirpes comerciais acima referidas (tabela 7). Contudo, estas estirpes comerciais, para além de partilharem grande parte das alterações cromossómicas (entre 68 a 70% das alterações eram partilhadas entre as duas), muitas destas pertenciam à mesma família de genes ou estavam associadas ao mesmo processo biológico (Tabela 9).

**Tabela 9. Exemplos de diferentes genes das mesmas famílias com alterações unicamente verificadas na estirpe ICV ou IOC.**

	<b>ICV</b>	<b>IOC</b>
<i>Hexose Transport</i>	<i>HXT8 e 9</i>	<i>HXT15 e 16</i>
<i>IMP Dehydrogenase</i>	<i>IMD2</i>	<i>IMD1</i>
<i>PHO</i> phate metabolism	<i>PHO11</i>	<i>PHO12</i>
<i>Pleiotropic Drug Resistance</i>	<i>PDR12</i>	<i>PDR3</i>

As restantes estirpes selvagens organizaram-se num outro ramo onde as alterações cromossómicas relativamente ao genoma de *S. cerevisiae* S288C foram maiores. Este ramo, constituído apenas por estirpes *S. cerevisiae* de origem vitivinícola dividiu-se em três subgrupos pouco relacionados entre si (correlação  $\sim 0$ ). As estirpes isoladas na Região Demarcada do Vinho Verde foram as mais díspares e por isso ficaram hierarquicamente mais afastadas das restantes embora tenham sido consideradas num mesmo subgrupo. Estas diferenças entre as duas estirpes provenientes da Região Demarcada do Vinho Verde eram esperadas com base num conjunto de características fenotípicas (resultados não publicados). A estirpe UM218 exibiu o cariótipo mais distinto de todas as estirpes selvagens (Apêndice 4). A maioria das estirpes apresentou um padrão maioritariamente de deleções, com poucas ou nenhuma regiões amplificadas. Contudo, na estirpe UM218 o aumento do número de alelos foi praticamente idêntico à deleção de alelos, relativamente à estirpe S288C (Tabela 6). As regiões onde se verificou acréscimo do número de alelos estavam, na sua maioria, localizadas fora dos habituais locais de perda de genes. Na extremidade direita do cromossoma XVI verificou-se uma exceção. UM218 ganhou um alelo de um gene que codifica uma glicogénio fosforilase, que foi deletado nas restantes estirpes de origem vitivinícola. Embora localizados em diferentes regiões cromossómicas, as ORFs com maior e menor número de cópias em *S. cerevisiae* UM218 distribuíram-se sensivelmente pelas mesmas categorias de funções moleculares. Esta coincidência, contudo, não é equilibrada no que se refere ao número de ORFs em cada categoria (Figura 15).



**Figura 15. Comparação da percentagem de ORFs da mesma categoria que em UM218 sofreram aumento e diminuição do número de cópias.**

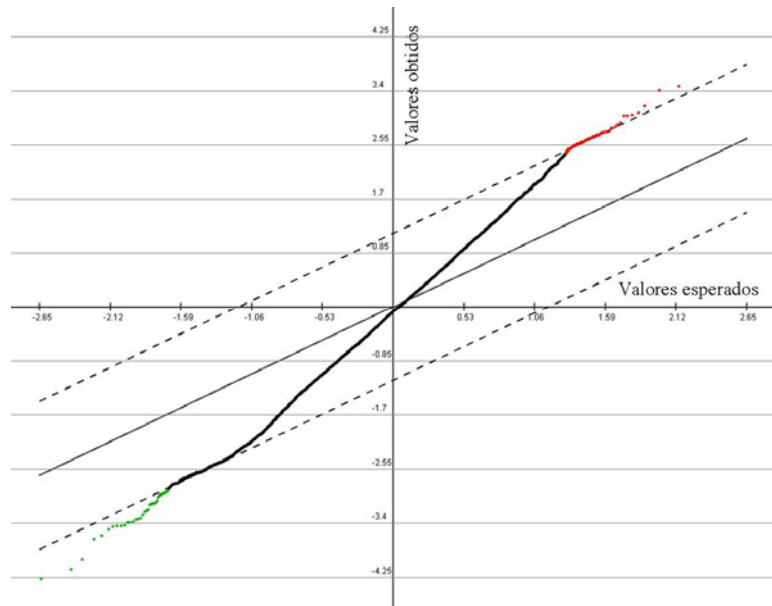
A vermelho está representada a percentagem de ORFs com número de cópias superior em UM218 relativamente a S288C. Analogamente, a verde estão as ORFs detectadas com menor número de cópias, daí o sinal “-” atribuído aos valores absolutos destas percentagens. A definição das categorias de funções moleculares, bem como o cálculo das percentagens relativas foram efectuadas pelo software FatiGO (Al-Shahrour et al., 2004; -, 2005; -, 2006).

Os subgrupos R1 e R2, quando analisados de forma semelhante ao efectuado para os subgrupos anteriormente descritos, revelaram-se semelhantes embora com algumas diferenças numa ou várias estirpes destes subgrupos. Contudo, a inspecção visual destas alterações mostrou-se demorada e sem resultados claros. Assim, para melhor caracterizar os subgrupos de estirpes fez-se uma análise sistematizada com software apropriado (ver resultados de SAM).

## 6. Significant Analysis of Microarrays (SAM)

Para tentar encontrar uma possível assinatura ou as características únicas e significativas de cada grupo de estirpes, relativamente a todas as outras, analisaram-se todas as estirpes selvagens em subgrupos, relativamente às restantes. Os subgrupos adoptados para esta análise foram escolhidos com base no agrupamento hierárquico obtido anteriormente. No caso das estirpes clínicas, a sua elevada semelhança e origem ecológica comum foram os critérios que levaram à definição do subgrupo J940. Apesar da baixa relação hierárquica entre as estirpes provenientes da Região Demarcada do Vinho Verde, estas apresentaram algumas semelhanças nos seus perfis cariotípicos (Apêndice 4) e uma vez que são as únicas deste nicho ecológico foram consideradas como subgrupo UM. As restantes estirpes, dos ramos definidos como R1 e R2, foram seleccionadas tendo em consideração o agrupamento hierárquico.

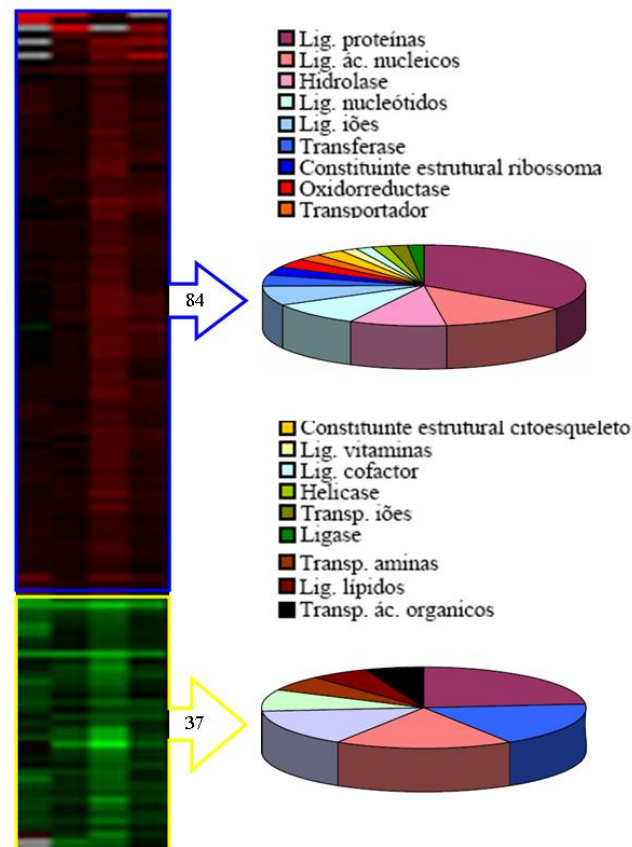
O padrão de variação dos valores de M (relativamente a S288C) dos genes de determinado subgrupo de estirpes foi comparado com o mesmo padrão para todas as restantes estirpes, gene a gene. Para esta análise, os parâmetros escolhidos foram, para cada subgrupo, um valor de  $\delta$  que correspondia ao maior número de genes com FRD (*False Discovery Rate*) = 0.  $\delta$  é a distância vertical entre a linha continua de declive 1 (valores observados = valores esperados) e a linha tracejada (declive  $1 \pm \delta$  unidades).



**Figura 16. Distribuição dos genes do subgrupo R2 relativamente às restantes estirpes de *S. cerevisiae* para análise SAM.**

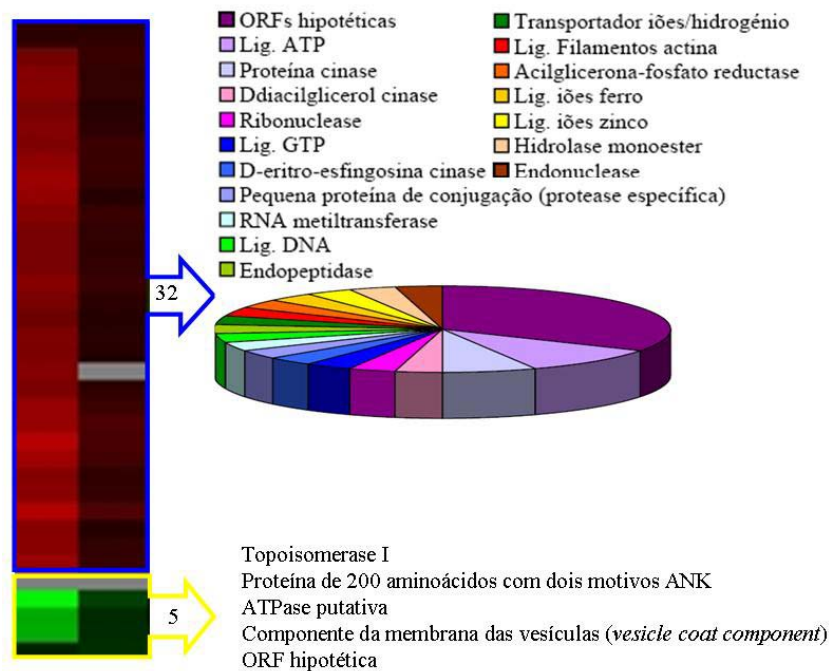
Os genes cujos valores de M estão representados a preto foram considerados como não significativos, os genes a vermelho e verde foram considerados significativamente positivos e negativos respectivamente.

Com este valor de FDR, a probabilidade dos genes encontrados terem sido escolhidos ao acaso (falsos positivos) foi nula. A Figura 17 exemplifica a representação gráfica dos valores de M (estirpe selvagem relativamente a S288C) dos genes detectados como significativamente diferentes num grupo de estirpes relativamente a todas as restantes. Na parte superior (rectângulo azul) encontram-se representados a vermelho os genes significativamente amplificados no sub-grupo de estirpes R2 comparativamente a todas as restantes estirpes. Na parte inferior (rectângulo amarelo) estão os genes com número significativamente menor número de cópias nas estirpes de R2 relativamente aos mesmos genes nas restantes estirpes. Da análise efectuada, com os critérios seleccionados, apenas se encontraram genes característicos dos subgrupos R2, J940 e UM.



**Figura 17. Representação gráfica dos genes significativamente diferentes nas estirpes autóctones de R2 relativamente às restantes estirpes (FDR = 0).**

As estirpes BB2453, 06L1FF11, 06L3FF02 e Fermol Rouge, pertencem ao subgrupo R2. A verde estão representados os genes com valores de M eram significativamente inferiores em R2 relativamente a todas as outras estirpes e a vermelho os genes cujos M são significativamente superiores em R2. Os genes cujos valores de M são representados a cinzento são inexistentes.



**Figura 18. Representação gráfica dos genes significativamente diferentes nas estirpes autóctones de UM relativamente às restantes estirpes (FDR = 0).**

As estirpes BB2453, 06L1FF11, 06L3FF02 e Fermol Rouge, pertencem ao subgrupo R2. A verde estão representados os genes com valores de M eram significativamente inferiores em R2 relativamente a todas as outras estirpes e a vermelho os genes cujos M são significativamente superiores em R2. Os genes cujos valores de M são representados a cinzento são inexistentes.

Finalmente, as estirpes clínicas, embora muito próximas hierarquicamente entre si, apenas apresentaram três genes significativamente diferentes de todas as restantes estirpes. *PLC1* (*PhosphoLipase C*), codifica uma fosfolipase C envolvida na formação de inositol 1,4,5-trifosfato, que é importante para o funcionamento do cinetocoro e diferenciação de pseudo-hifas. A invertase codificada pelo gene *SUC2* (*SUCrose fermentation*) que hidrolisa sucrose e é regulada na forma glicosilada pelo mecanismo de repressão da glucose. Finalmente, o gene *SOR1* (ou *SDH1*, *Sorbitol DeHydrogenase*) da sorbitol desidrogenase que é induzido na presença deste soluto metabolizado.

No caso do subgrupo R1, mesmo fazendo uma nova análise SAM a pequenos subgrupos, não foram encontrados genes significativamente diferentes nestas estirpes. Assim, IOC, ICV, BB1235, 06L3FF15, 06L6FF20 e Davis Lalvin5222 não apresentaram qualquer característica que as distinguisse significativamente das restantes.

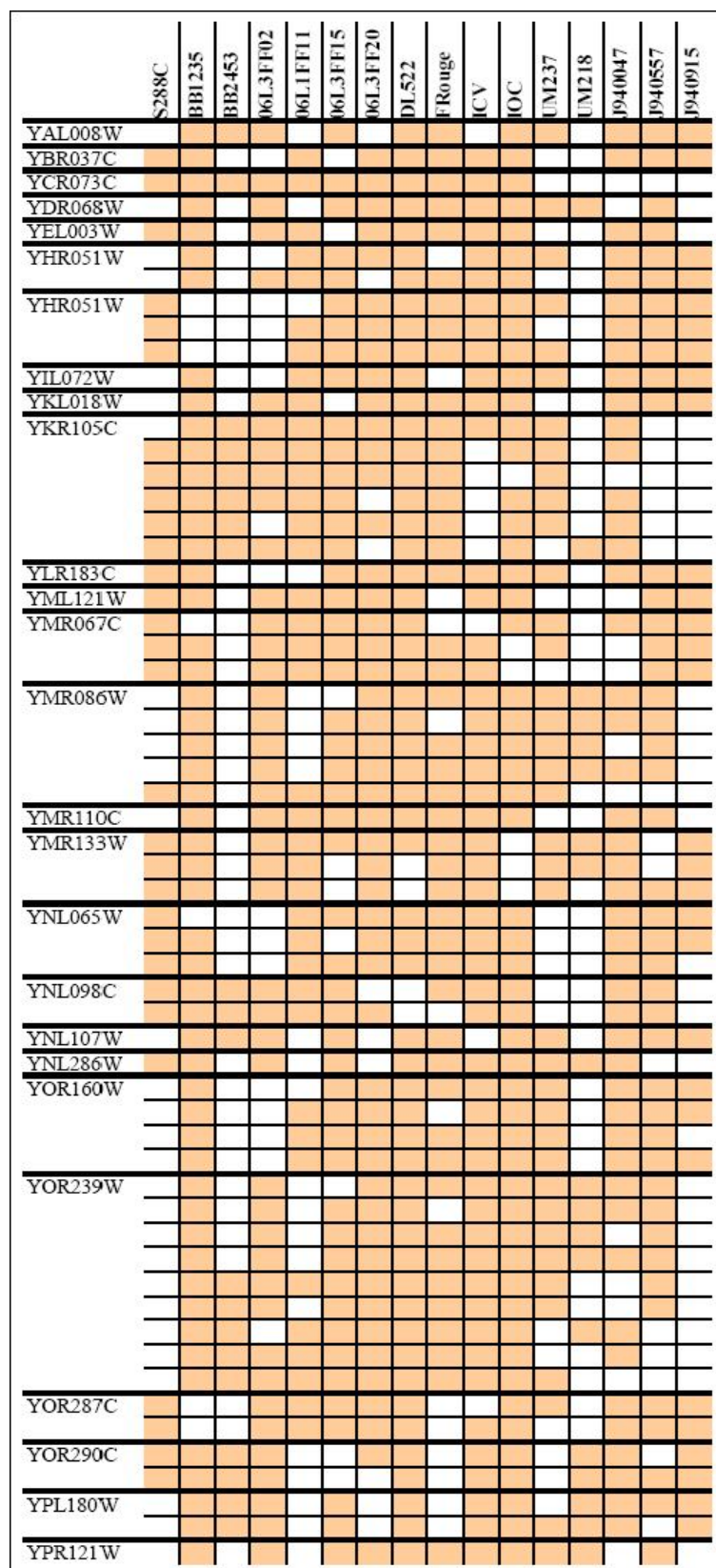
## 7. *Bi-clustering*

*Bi-clustering* é uma técnica de agrupamento de dados local que analisa simultaneamente as duas dimensões da matriz de dados, as linhas e as colunas. Assim, ao contrário do agrupamento hierárquico obtido anteriormente, com técnicas de *bi-clustering* os genes e as amostras podem ser agrupados simultaneamente de forma local e não global. O objectivo do *bi-clustering* é identificar subgrupos de genes com comportamento semelhante em determinado subgrupo de amostras. Acrescenta-se, ainda, que neste caso de agrupamento de dados, um gene/amostra pode pertencer a mais do que um grupo ou mesmo a nenhum (Madeira & Oliveira, 2004).

O algoritmo de cálculo dos *bi-clusters* utilizado foi o xMotif (Murali & Kasif, 2003). Este algoritmo baseia-se na procura de bi-clusters com evolução coerente independentemente dos valores numéricos exactos. Assim, genes com o mesmo padrão de deleção ou amplificação foram agrupados em conjunto com as amostras nas quais manifestaram o mesmo comportamento – motivos conservados de genes através de categorias (amostras).

As ORFs não anotadas foram eliminadas desta análise. Cada linha de quadriculas laranjas e brancas representa um *bi-cluster* onde a linha identifica o gene e as colunas identificam as amostras (coloridas a laranja estão as amostras que pertencem a cada *bi-cluster*). Nesta análise pôde observar-se que o mesmo gene pôde agrupar as mesmas amostras de forma diversa consoante uma ou mais amostras entra ou sai do *bi-cluster*. Por exemplo, YOR290C (*SNF2*) é responsável pela formação de dois *bi-clusters* diferentes em apenas 2 estirpes, quando 06L3FF15 faz parte do *bi-cluster*, J940557 não, e vice-versa.





**Figura 19. Representação gráfica do resultado do *bi-clustering* (xMotifs) de amostras e genes.**

Cada linha representa um *bi-cluster* no qual a ORF definida à esquerda teve comportamento conservado nas amostras (coloridas a laranja).

**Tabela 10. Nomes de genes e respectivas funções das ORFs obtidas no *bi-clustering*.**

YAL008W	<i>FUN14</i> ( <b>F</b> unction <b>U</b> nknown <b>N</b> ow) Proteína mitocondrial de função desconhecida
YBR037C	<i>SCO1</i> ( <b>S</b> upressor of <b>C</b> ytochrome <b>O</b> xidase <b>d</b> eficiency) Proteína da membrana interna da mitocondria de ligação ao cobre, necessária a respiração
YCR073C	<i>SSK22</i> ( <b>S</b> upressor of <b>S</b> ensor <b>K</b> inase) MAP cinase cinase cinase da via de sinalização HOG1
YDR068W	<i>DOS2</i> Proteína de função desconhecida
YEL003W	<i>GIM4</i> ( <b>G</b> ene <b>I</b> nvolved in <b>M</b> icrotubule <b>b</b> iogenesis) Subunidade de complexo co-chaperone <i>prefolding</i> , ligação específica e chaperonina citosólica
YHR051W	<i>COX6</i> (Cytochrome c <b>O</b> Xidase) Subunidade VI de citocromo c oxidase, ultimo elemento da cadeia transportadora de electrões da membrana mitocondrial (regulado por níveis de oxigénio)
YHR051W	<i>DSE1</i> ( <b>D</b> aughter <b>S</b> pecific <b>E</b> xpression) Proteína específica de células-filhas, afecta divisão celular e sensibilidade da parede celular a drogas
YIL072W	<i>HOP1</i> ( <b>H</b> omolog <b>P</b> airing) Proteína de ligação ao DNA específica de meiose
YKL018W	<i>SWD2</i> ( <b>S</b> et1c, <b>W</b> D40 repeat protein) Subunidade de complexo COMPASS (Set1C) que metila histona H3 e está envolvido no silenciamento de telómeros
YKR105C	Transportador putativo da Major Facilitator Superfamily
YLR183C	<i>TOS4</i> (Target of SBF) Factor transição especialmente associado a promotores de genes envolvidos na resposta a feromonas e ciclo celular
YML121W	<i>GTR1</i> ( <b>G</b> TP binding protein <b>R</b> esemblance) Proteína citoplasmática de ligação ao GTP, envolvida no transporte de fosfato
YMR067C	<i>UBX4</i> ( <b>U</b> biquitin regulatory <b>X</b> ) Dominio com proteína que interage com Cdc48p
YMR086W	Proteína de função desconhecida que interage com ribossomas
YMR110C	<i>HFD1</i> ( <b>H</b> omolog of <b>F</b> atty aldehyde <b>D</b> ehydrogenase) Desidrogenase de aldeídos gordos localizados na membrana externa da mitocondria
YMR133W	<i>REC114</i> (RECombination) Proteína envolvida em estadios iniciais de recombinação meiotica
YNL065W	<i>AQR1</i> ( <b>A</b> -acids, azoles- <b>Q</b> -quinidine, quinine- Resistance) Transportador de membrana de major facilitator superfamily, confere resistência a ácidos monocarboxílicos de cadeia curta
YNL098C	<i>RAS2</i> (Homologous to RAS proto-oncogene) Proteína de ligação a GTP reguladora de resposta a défice de azoto, esporulação e crescimento filamentoso
YNL107W	<i>YAF9</i> (Yeast homolog of human leukemogenic protein AF9) Subunidade de histona H4 acetil transferase, pode antagonizar o silenciamento proximo de telómeros
YNL286W	<i>CUS2</i> ( <b>C</b> old sensitive <b>U</b> 2 snRNA <b>S</b> upressor) Proteína de ligação a snRNA U2 e Prp11p
YOR160W	<i>MTR10</i> (MRNA Transport defective) Receptor de transporte para interior do núcleo
YOR239W	<i>ABP140</i> (Actin Binding Protein) Proteína não essencial de ligação a filamentos de actina
YOR287C	Pode estar envolvido na síntese de rRNA e ribossomas
YOR290C	<i>SNF2</i> ( <b>S</b> ucrose <b>N</b> on <b>F</b> ermenting) Subunidade catalítica de complexo de remodelação da cromatina SWI/SNF envolvido na regulação da transcrição
YPL180W	<i>TCO89</i> ( <b>T</b> or <b>C</b> omplex <b>O</b> ne) Subunidade de complexo Tor que regula crescimento em resposta a disponibilidade nutricional
YPR121W	<i>THI22</i> ( <b>T</b> Hiamine metabolism) Proteína semelhante a hidroximetilpirimidina fosfato cinase, não fundamental à biossíntese de tiamina

## **8. Assinaturas de leveduras vínicas: uma comparação**

Dunn e colaboradores (2005) identificaram uma assinatura característica de leveduras vínicas comerciais. Para comprovar a existência desta assinatura comparou-se a lista descrita por estes autores com os consensos obtidos para as quatro estirpes comerciais utilizadas neste estudo e para todas as estirpes selvagens de origem vitivinícola. Embora o desenho experimental tenha sido diferente, a estirpe S288C foi utilizada como referência, e o tratamento de dados foi semelhante em ambos os estudos. Assim, embora comparações quantitativas não tenham sido efectuadas, devido ao diferente número de hibridações por amostra, à não realização de réplicas técnicas e a uma diferente colecção de sondas impressas nos slides, fizeram-se comparações qualitativas. A tabela 8 sumaria a comparação das assinaturas encontradas.

Dos 49 genes alterados no consenso de Dunn e colaboradores (2005), 29 são comuns ao consenso obtido com as 4 estirpes comerciais neste estudo, dos quais 27 são comuns ao consenso alcançado com todas as estirpes de origem vitivinícola (estirpes da Região Demarcada da Bairrada, estirpes Região Demarcada do Vinho Verde e estirpes comerciais). Todos estes genes comuns à “assinatura de leveduras vínicas” foram detectados com inferior número de cópias em todas as estirpes selvagens comparativamente a S288C.

Outras das alterações definidas pelos mesmos autores como características de todas as leveduras vínicas não foram detectadas em todas as estirpes deste estudo: as amplificações no terminal esquerdo do cromossoma III não deverão ser características das leveduras vínicas mas sim de algumas leveduras vínicas.

Os transportadores de açúcares foram excluídos destas listas, uma vez que a variação na utilização de um conjunto específico de transportadores depende do tipo de substrato utilizado, e consequentemente, do tipo de ambiente colonizado.

**Tabela 11. Características da assinatura de leveduras vínicas [Dunn *et al.* (2005)] presentes nas leveduras autóctones deste estudo.**

Entre parêntesis está o número total de genes encontrados. X significa que a respectiva alteração cromossômica descrita por Dunn e colaboradores (2005) foi encontrada nas estirpes deste estudo.

Nome da ORF	Nome do gene	Estirpes comerciais (281)	Todas estirpes vínicas (202)
<b>Maior número de cópias</b>			
YAR066W		-	-
YAR068W		-	-
YAR069C		-	-
YAR070C		-	-
YAR071W	<i>PHO11</i>	-	-
YAR073W	<i>IMD1</i>	-	-
YAR075W		-	-
YPL276W		-	-
YPL275W		-	-
YPL274W	<i>FDH2</i>	-	-
YPL273W	<i>SAM3</i>	-	-
YPL272C	<i>SAM4</i>	-	-
<b>Menor número de cópias</b>			
YAR007C	<i>RFA1</i>	X	X
YAR008W	<i>SEN34</i>	X	X
YAR014C	<i>BUD14</i>	X	X
YAR015W	<i>ADE1</i>	X	X
YBL007C	<i>SLA1</i>	-	-
YBL006C	<i>LDB7</i>	X	X
YBL005W	<i>PDR3</i>	X	X
YBR013C		X	X
YHR052W	<i>CIC1</i>	-	-
YHR053C	<i>CUP1-1</i>	-	-
YHR054C		-	-
YHR055C	<i>CUP1-2</i>	-	-
YHR056C	<i>RSC30</i>	-	-
YJR024C	<i>MDE1</i>	X	X
YJR025C	<i>BNA1</i>	X	X
YJR026W	(Ty)	X	X
YJR030C		X	X
YJR031C	<i>GEA1</i>	X	X
YJR032W	<i>CPR7</i>	-	-
YJR033C	<i>RAV1</i>	-	-
YML047C	<i>PRM6</i>	X	X
YML046W	<i>PRP39</i>	X	X
YML043C	<i>RRN11</i>	X	X
YML042W	<i>CAT2</i>	X	X
YML041C	<i>VPS71</i>	X	X
YML038C	<i>YMD8</i>	X	X
YML037C		X	X
YMR043W	<i>MCM1</i>	X	X
YMR044W	<i>IOC4</i>	X	X
YMR046W-A		X	X
YMR047C	<i>NUP116</i>	X	X
YMR048W	<i>CSM3</i>	X	X
YMR049C	<i>ERB1</i>	X	X
YMR052W	<i>FAR3</i>	X	X
YMR052C-A		X	X
YOL165C	<i>AAD15</i>	X	-
YOL164W		X	-

O cromossoma I é o mais variável do genoma de *S. cerevisiae*, pelo que eram esperadas alterações significativas. Neste, os genes *RFA1* (**R**eplication **F**actor **A**), *SEN34* (**S**plicing **E**ndonuclease), *BUD14* (**BUD** site selection) e *ADE1* (**ADE**nine requiring) foram detectados com número de alelos inferior em todas as estirpes selvagens relativamente a S288C.

**Tabela 12. Assinatura de leveduras vínicas, cromossoma I**

<b>Cromossoma I</b>		
	<b>Função descrita</b>	<b>Possível alteração fenotípica</b>
<i>RFA1</i>	Factor de replicação A, parte do complexo responsável pela verificação de danos no DNA	
<i>SEN34</i>	Codifica uma das quatro subunidades da endonuclease envolvida no <i>splicing</i> de tRNAs	Possível inactivação da clivagem do terminal 5' do local de <i>splicing</i> do intrão
<i>BUD14</i>	Reguladora de Glc7p, subunidade catalítica de uma fosfatase responsável pela interacção dos microtúbulos com os cromossomas segregados em metáfase	Possíveis irregularidades na gemulação
<i>ADE1</i>	Um dos nove genes da família <i>ADE</i> responsáveis pela biossíntese <i>de novo</i> de purinas	As restantes cópias de <i>ADE</i> poderão estar a assegurar este metabolismo

No cromossoma II, uma ORF ainda não anotada, bem como os genes *LDB7* (**L**ow **D**ye **B**inding) e *PDR3* (**P**leitropic **D**rug **R**esistance) são comuns a todas as assinaturas. Tal como no estudo de Dunn e colaboradores (2005), no cromossoma X foi encontrada uma ORF não anotada, um transposão, *MDE1* (**M**ethylthio-**ri**bulose-**1**-**phosphate DE**hydratase), *BNA1* (**B**iosynthesis of **N**icotinic **A**cid) e *GEA1* (**G**uanine nucleotide **E**xchange on **ARF**). *MDE1* é nome dado pela SGD ao gene da ORF YJR024C.

**Tabela 13. Assinatura de leveduras vínicas, cromossoma II**

<b>Cromossoma II</b>		
	<b>Função descrita</b>	<b>Possível alteração fenotípica</b>
<i>LDB7</i>	Componente do complexo RSC (remodelação da estrutura da cromatina), interage com Rsc3p e outras proteínas	
<i>PDR3</i>	Codifica um factor de transcrição com motivo <i>zinc-finger</i> associado à resistência múltipla a drogas; regula a expressão de transportadores dependentes de ATP	Tanto vinhas como lagares estão habitualmente sujeitos à adição de produtos químicos (tratamento de pragas ou controle de qualidade da fermentação). Constante pressão selectiva efectuada por estes agentes poderá ter levado à fixação de características favoráveis à sobrevivência nestes ambientes

No cromossoma X foi encontrada em comum com Dunn e colaboradores (2005) uma ORF não caracterizada, um transposição, *MDE1* (*Methylthio-ribulose-1-phosphate Dehydratase*), *BNA1* (*Biosynthesis of Nicotinic Acid*) e *GEA1* (*Guanine nucleotide Exchange on ARF*).

**Tabela 14. Assinatura de leveduras vínicas, cromossoma X**

<b>Cromossoma X</b>		
	<b>Função descrita</b>	<b>Possível alteração fenotípica</b>
<i>MDE1</i>	Desidratase envolvida na via de recuperação de metionina ( <i>methionine salvage pathway</i> )	Poderá induzir resistência a caspofungina (anti fúngico com alvo na parede celular)
<i>BNA1</i>	Enzima necessária para biossíntese de ácido nicotínico (via de degradação aeróbia do triptofano)	
<i>GEA1</i>	Proteínas necessárias à activação das Arf, envolvidas no transporte intracelular entre Golgi e o Reticulo (Chavier & Goud, 1999)	Deleção não inviabilizante para a célula (Peyroche <i>et al.</i> , 1996)

Finalmente no cromossoma XIII, o maior dos cromossomas onde foi encontrada uma parte da “assinatura de leveduras vínicas”, dois grandes segmentos estão diminuídos relativamente a S288C. *PRM6* (*Pheromone-Regulated Membrane*), *PRP39* (*Pre-mRNA Processing*), *RRN11* (*Regulation of RNA polymerase I*), *CAT2* (*Carnitine AcetylTransferase*), *VPS71* (*Vacuolar Protein Sorting*), *YMD8* e uma ORF não caracterizada formam um bloco. *MCMI* (*MiniChromosome Maintenance*), *IOC4* (*Iswi One Complex*), *NUP116* (*NUclear Pore*), *CSM3* (*Chromosome Segregation in Meiosis*), *ERB1*, *FAR3* (*Factor ARrest*) e duas ORFs duvias constituem o segundo bloco.

**Tabela 15. Assinatura de leveduras vínicas, cromossoma XIII**

<b>Cromossoma XIII</b>		
	<b>Função descrita</b>	<b>Possível alteração fenotípica</b>
<i>PRM6</i>	Proteína membrana envolvida no último passo de fusão celular (processo de reprodução sexuada)	Estas proteínas são codificadas por vários genes, reprodução sexuada não deverá estar comprometida
<i>PRP39</i>	<i>Splicing</i> de pré-mRNA.	Possíveis erros no <i>splicing</i> de pré-mRNAs
<i>RRN11</i>	Componente de factor de transcrição de rDNA	Síntese proteica poderá estar de algum modo comprometida
<i>CAT2</i>	Responsável pela transferase de grupos acetil activados para a carnitina	Carnitina acetiltransferases são fundamentais na $\beta$ -oxidação de ácidos gordos (peroxissomas), este processo metabólico poderá estar afectado
<i>VPS71</i>	Envolvida no tráfego de vesículas	Proteínas de vacuolos podem ter o seu tráfego alterado
<i>YMD8</i>	Transportador semelhante ao Vrg4p (transportador essencial à glicosilação de proteínas e lípidos no Complexo de Golgi)	Vrg4p cujo gene <i>VRG4</i> ( <i>Vandate Resistance Glycosylation</i> ) permanece inalterado assegura glicosilação
<i>MCM1</i>	Determinação do tipo de célula, metabolismo de arginina, resposta a feromonas, reprodução sexuada, progressão do ciclo celular e iniciação da replicação	Sensibilização a elevados níveis salinos (Abraham & Vershon, 2005)
<i>IOC4</i>	Em conjunto com Isw1p e Ioc2p formam um complexo (Isw1b) com actividade ATPase que coordena a elongação e terminação da transcrição	Transcrição poderá ser afectada
<i>NUP116</i>	Subunidade do complexo do poro nuclear (NPC)	Sensibilidade à temperatura e problemas na exportação de tRNA
<i>CSM3</i>	Necessário à correcta segregação de cromossomas durante a meiose	Possíveis alterações na meiose
<i>ERB1</i>	Necessário à maturação de rRNA (25S e 5,8S)	Não só as referidas 25S e 5,8S poderão deixar de ser produzidas como também o processamento de pré-rRNA 35S poderá ser afectado
<i>FAR3</i>	Regula ciclo celular em resposta a feromonas, previne recuperação prematura do ciclo celular após paragem por resposta a sinais de reprodução sexuada	Resposta do ciclo celular a feromonas não estará de todo alterada uma vez que outros genes envolvidos neste processo permanecem nos genomas estudados





## **DISCUSSÃO GERAL**

Para todas as estirpes analisadas nunca se verificou uma total deleção ou amplificação de um ou vários cromossomas. Embora alguns estudos (Guijo *et al.*, 1997; Miklos *et al.*, 1997) tenham referido a alteração da ploidia como um evento comum em leveduras vínicas, nas estirpes vínicas estudadas não foi detectada nenhuma ocorrência de aneuploidia ou poliploidia.

Analisando o agrupamento hierárquico de todas as estirpes e tendo em consideração os perfis de restrição do locus *MET2* (Anexo 1) efectuado para todas as estirpes selvagens com excepção das estirpes provenientes da Região Demarcada do Vinho Verde, pôde inferir-se que estas eram estirpes de *S. cerevisiae*. A variabilidade genética de todas as estirpes autóctones estudadas, visível nos cariótipos gerados pelo CGH-Miner (Apêndice 4), foi grande mesmo tratando-se de uma só espécie. Esta diferença genética reflecte-se nos diferentes nichos ecológicos aos quais cada estirpe está mais adaptada. As três estirpes clínicas são semelhantes entre si mas apresentaram importantes diferenças relativamente às estirpes vínicas. Estas estirpes J940557 e J940915 estavam bastante próximas, o que era de certa forma esperado uma vez que os perfis inter- $\delta$  destas duas estirpes foram idênticos (Anexo 2). As duas estirpes da Região Demarcada do Vinho Verde também apresentaram diferenças relativamente às outras. A estirpe UM218 foi a que apresentou mais diferenças com números equivalentes de perda e ganho de alelos. Estas diferenças reflectiram-se no posicionamento relativo das referidas estirpes no agrupamento hierárquico.

Uma vez que as estirpes autóctones estudadas provêm de ambientes muito diferentes, as 14 ORFs detectadas em menor número de cópias em todas as estirpes, relativamente a S288C, não poderão ser associados a nenhum nicho ecológico.

No consenso das estirpes selvagens, a maioria das alterações cromossómicas detectadas estava organizada em processos biológicos e funções moleculares que apresentam alguma redundância. Esta redundância está associada a famílias de genes ou genes homólogos e a processos biológicos, sugerindo que a perda de alguns alelos não terá um impacto significativo no fenótipo porque a sua função pode ser compensada por outros genes. Por outro lado, a reduzida perda de cópias de alelos nas categorias conservadas suporta a hipótese de que as funções biológicas fundamentais para a célula são pouco afectadas pela variabilidade genética observada entre estirpes. Por exemplo, genes envolvidos na

autofagia, crescimento filamentoso, segregação cromossómica ou desenvolvimento de estruturas anatómicas não são muito afectados.

Após a análise do consenso de todas as estirpes, tentaram encontrar-se, para cada subgrupo de estirpes, as características únicas responsáveis pelo seu agrupamento. Os resultados até aqui obtidos tinham-se baseado na detecção do maior ou menor número de cópias definido pelo CGH-Miner. Uma vez que este software construiu todas as representações gráficas com base numa janela de genes, estas alterações não reflectiram qualquer alteração isolada de um único gene. A SAM é diferente da análise efectuada com o CGH-Miner, pois naquela um grupo de genes cuja variação de valores de M é significativamente diferente num determinado grupo de estirpes relativamente às restantes é detectado independentemente deste formar um bloco de genes adjacentes ou apenas genes isolados.

Como referido nos Resultados, apenas os subgrupos UM, J940 e R2 apresentaram genes significativamente diferentes das restantes estirpes. Estes genes, característicos das estirpes que formaram cada subgrupo, deverão ser responsáveis pelo agrupamento destas nos subgrupos definidos. No caso do subgrupo R1, mesmo fazendo uma nova análise a pequenos subgrupos, com maior correlação hierárquica entre as estirpes, não foram encontrados genes significativamente diferentes. Duas possíveis explicações poderão estar na origem desta ausência de diferenças: por um lado, as estirpes de determinado subgrupo poderão ser de tal forma diferentes entre si que a variabilidade entre subgrupos será insignificante; por outro lado, as estirpes de determinado subgrupo poderão ser tão semelhantes entre si como relativamente às restantes, o que impossibilitaria a distinção de genes significativamente diferentes. Neste ultimo caso, a estabilidade genética da espécie sobrepor-se-ia a variabilidade das estirpes.

O *bi-clustering* analisa simultaneamente genes e amostras permitindo agrupar estirpes de acordo com categorias de genes. Diferentes genes poderão agrupar as mesmas amostras de formas diferentes consoante uma ou mais amostras são adicionadas. A comparação dos resultados de CGH-Miner, SAM e *bi-clustering* foi difícil de fazer pois os métodos de análise são diferentes, sendo por isso justificável a quase completa ausência de resultados comuns. Nenhum dos métodos poderia ser substituído por outro, devendo ser usados para diferentes fins e com diferentes interpretações. CGH-Miner será talvez o mais indicado para análise exclusiva de alterações cromossómicas relativamente a uma referência. SAM

poderá ser útil no cruzamento de informações de aCGH e expressão genética. *Bi-clustering* será apropriado para agrupamento de estirpes, ou no caso de experiências de expressão genética agrupamento de condições de crescimento ou tempos, tendo em consideração as classes de genes igualmente afectadas.

Da comparação das “assinaturas de leveduras vínicas” utilizadas neste estudo e da obtida por Dunn e colaboradores (2005) observou-se que, embora alguns genes tenham sido igualmente alterados nos genomas de todas as estirpes, as alterações cromossómicas características deste tipo de leveduras não foram todas, nem exclusivamente, as apontadas por aqueles autores. Em conjunto com o verificado pelos mesmos, a perda de fragmentos cromossómicos foi mais frequente que o ganho destes em todas as estirpes com excepção de UM218. Contudo, nestes estudos apenas se puderam detectar alterações de número de cópias de genes de *S. cerevisiae* S288C. Nada se pôde concluir quanto ao DNA genómico com origem noutras espécies de *Saccharomyces* (*S. cerevisiae* ou não), uma vez que as sondas impressas nos slides apenas cobriam o ORFeoma da estirpe *S. cerevisiae* S288C.

## **CONCLUSÕES**

As estirpes *S. cerevisiae* autóctones analisadas neste estudo não puderam, com base nas alterações cromossómicas detectadas, ser agrupadas por nicho ecológico uma vez que mesmo apresentando padrões de alterações cromossómicas comuns, estas alterações não estão relacionadas com a sua proveniência.

Praticamente todas as alterações detectadas nas estirpes autóctones deste estudo corresponderam a deleções ou decréscimos no número de cópias de alelos e poucas foram as amplificações de fragmentos cromossómicos. Mesmo com diferentes proveniências, todas as quinze estirpes autóctones estudadas partilharam algumas diferenças cromossómicas relativamente à estirpe de referência S288C: *ASP3*, *ACS2*, *RNH203* e *MAS1*. Outros genes, como os envolvidos na recombinação de segmentos teloméricos, transposões, transportadores ou permeases, sensores de drogas ou stresses, factores de transcrição ou tradução, foram alguns dos mais frequentemente alterados.

No que se refere ao ganho de material genético com origem noutras estirpes de *Saccharomyces*, com as quais as estirpes estudadas possam interagir nos seus habitats, nada se pôde concluir uma vez que as sondas utilizadas apenas correspondem ao ORFeoma de *S. cerevisiae* S288C. Também rearranjos resultantes de troca de posição de segmentos cromossómicos não puderam ser detectados com esta técnica.

Uma assinatura característica foi apontada para alguns grupos de estirpes mas não para todas. Uma vez que tanto as vinhas como os lagares de fermentação são habitualmente sujeitos à adição de produtos químicos, para tratamento de pragas ou controle de qualidade da fermentação, e outros stresses, a constante pressão selectiva terá levado à fixação das características observadas por serem favoráveis à sobrevivência nestes ambientes. As estirpes clínicas, sob diferentes pressões selectivas, terão maior resistência a outros factores ambientais e por isso estão mais adaptadas aos ambientes que habitualmente colonizam. Assim, o tempo e o número de gerações permitiram fixar características genómicas responsáveis por variações nos fenótipos destas estirpes *S. cerevisiae*.

Finalmente, este estudo levantou novas questões e indicou novas estratégias de investigação. RT-PCR e sequenciação de DNA deverão ser usados para confirmar os resultados de aCGH. A utilização de *tiled arrays* de alta densidade poderia ser útil para identificar SNPs a nível global. Uma refinação dos resultados seria esperada se mais

estirpes de *S. cerevisiae* e de outras espécies próximas desta fossem adicionadas ao estudo. Adicionalmente, estudos de expressão genética das mesmas estirpes em condições de fermentação específicas poderão dar uma melhor indicação dos efeitos da diversidade genética no fenoma e na adaptação ambiental.





## **REFERÊNCIAS**

- Abraham, D.S., Vershon, A.K. (2005) *N-terminal arm of Mcm1 is required for transcription of a subset of genes involved in maintenance of the cell wall*. Eukaryot Cell 4(11): 1808-19
- Al-Shahrour, F., Díaz-Uriarte, R. & Dopazo, J. (2004) *FatiGO: a web tool for finding significant associations of Gene Ontology terms with groups of genes*. Bioinformatics 20: 578-580
- Al-Shahrour, F., Minguéz, P., Tárraga, J., Montaner, D., Alloza, E., Vaquerizas, J.M.M., Conde, L., Blaschke, C., Vera, J. & Dopazo, J. (2006) *BABELOMICS: a systems biology perspective in the functional annotation of genome-scale experiments*. Nucleic Acids Research 34: 472-476
- Al-Shahrour, F., Minguéz, P., Vaquerizas, J.M., Conde, L. & Dopazo, J. (2005) *Babelomics: a suite of web-tools for functional annotation and analysis of group of genes in high-throughput experiments*. Nucleic Acids Research 33: 460-464
- Bakalinsky, A.T., Snow, R. (1990) *Conversion of wine strains of Saccharomyces cerevisiae to Heterothallism*. Appl Environ Microbiol 56 (4):849-857
- Bakalinsky, A.T., Snow, R. (1990) *The chromosomal constitution of wine strains of Saccharomyces cerevisiae*. Yeast 6 (5): 367-82
- Barrett, M.T., Scheffer, A., Bem-Dor, A., Sampas, N., Lipson, D., Kincaid, R., Tsang, P., Curry, B., Baird, K., Meltzer, P.S., Yakhini, Z., Bruhn, L., Laderman, S. (2004) *Comparative genomic hybridization using oligonucleotide microarrays and total genomic DNA*. Proc Natl Acad Sci USA 101 (51): 17765-70
- Beltran, G., Torija, M.J., Novo, M., Ferrer, N., Poblet, M., Guillamon, J.M., Rozes, N., Mas, A (2002) *Analysis of yeast populations during alcohol fermentation: a six year follow-up study*. Syst Appl Microbiol 25 (2): 287-93
- Benitez, T., Martinez, P., Codon, A.C. (1996) *Genetic constitution of industrial yeast*. Microbiologia 12 (3): 371-84

- Bon, E.P. *et al.* (1997) *Asparaginase II of Saccharomyces cerevisiae. GLN3/URE2 regulation of a periplasmic enzyme.* Appl Biochem Biotechnol 63-65:203-12
- Bouza, E., Munoz, P. (2004) *Saccharomyces cerevisiae: the end of the innocence.* Rev Esp Quimioter 17 (3): 227-31
- Brookfield, J. (1992) *Can genes be truly redundant?* Curr Biol. 2 (10): 553-4
- Cavalieri, D., Townsend, J.P., Hartl, D.L. (2000) *Mainfold anomalies in gene expression in a vineyard isolate of Saccharomyces cerevisiae revealed by DNA microarray analysis.* Proc Natl Acad Sci USA 97 (22):12369-74
- Cech, T.R., Nakamura, T.M., Lingner, J. (1997) *Telomerase is a true reverse transcriptase. A review.* Biochemistry (Mosc.) 62: 1202-1205
- Chalker, D.L. & Sandmeyer, S.B. (1992) *Ty3 integrates within the region of RNA polymerase III transcription initiation.* Genes Dev 6(1):117-28
- Chan, C.S.M., Tye, B.K. (1983) *Organization of DNA sequences and replication origins at yeast telomeres.* Cell 33 (2): 563-73
- Chavrier, P., Goud, B. (1999) *The role of ARF and Rab GTPases in membrane transport.* Curr Opin Cell Biol 11(4): 466-75
- Cherry, J.M., Adler, C., Ball, C., Chervitz, S.A., Dwight, S.S., Hester, E.T., Jia, Y., Juvik, G., Roe, T., Schroeder, M., Weng, S., Botstein, D. (1998) *SGD: Saccharomyces Genome Database.* Nucleic Acids Res 26(1): 73-9
- Chertow, G.M., Marcantonio, E.R., Wells, R.G. (1991) *Saccharomyces cerevisiae empyema in a patient with esophago-pleural fistula complicating variceal sclerotherapy.* Chest. 99 (6): 1518-9
- Clemons, K.V., McCusker, J.H., Davis, R.W., Stevens, D.A. (1994) *Comparative pathogenesis of clinical and nonclinical isolates of Saccharomyces cerevisiae.* J. Infect Dis 169 (4):859-67

- Codon, A.C., Benitez, T., Korhola, M. (1998) *Chromosomal polymorphism and adaptation to specific industrial environments of Saccharomyces strains*. Appl Microbiol Biotechnol 49 (2): 154-63
- Codon, A.C., Gasent-Ramirez, J.M., Benitez, T. (1995) *Factors which affect the frequency of sporulation and tetrad formation in saccharomyces cerevisiae baker's yeast*. Appl Environ Microbiol. 61 (4): 1677
- Cooper, M.G. (2000) *The Cell: A molecular Approach*. 2<sup>nd</sup> Ed. Massachusetts. Sinauer Associates, Inc.
- Culver, G.M., McCraith, S.M., Consual, S.A., Stanford, D.R., Phizicky, E.M. (1997) *A 2'-phosphotransferase implicated intrRNA splicing is essential in Saccharomyces cerevisiae*. J Biol Chem 272 (20): 13203-10
- Darwin C. (1859) *On the origin f species by means of natural selection or the preservation of favoured races in the struggle for life*. Edn.Penguin, London
- Demuyter, C., Lollier, M., Legras, J.L., Le Jeune, C. (2004) *Predominance of Saccharomyces uvarum during spontaneous alcoholic fermentation, for three consecutive years, in an Alsatian winery*. J Appl Microbiol 97 (6): 1140-8
- Devine, S.E. & Boeke, J.D. (1996) *Integration of the yeast retrotransposon TyI is targeted to regions upstream of genes transcribed by RNA polymerase III*. Genes Dev 10(5):620-33
- Dilda, P.J., Don, A.S., Tanabe, K.M., Higgins, V.J., Allen, J.D., Dawes, I.W., Hogg, P.J. (2005) *Mechanism of selectivity of an angiogenesis inhibitor from screening a genome-wide set of Saccharomyces cerevisiae deletion strains*. J Natl Cancer Inst 97(20): 1539-47
- Draghici, S. (2003) *Cluster Analysis in Data analysis tools for DNA Microarrays*, New York, Chapman and Hall/CRC
- Dujon B. (1996) *The yeast genome project: what did we learn?* Trends Genet. 12 (7): 263-70

Dujon B., Sherman D., Fischer G., Durrens P., Casaregola S., Lafontaine I., Montigny J., Marck C., Neuvéglise C., Talla, E., Goffard, N., Frangeul, L., Aigle, M., Anthouard, V., Babour, A., Barbe, V., Barnay, S., Blanchin, S., Beckerich, J.M., Beyne, E., Bleykasten, C., Boisramé, A., Boyer, J., Cattolico, L., Confanioleri, F., Daruvar, A., Despons, L., Fabre, E., Fairhead, C., Ferry-Dumazet, H., Groppi, A., Hantraye, F., Hennequin, C., Jauniaux, N., Joyet, P., Kachouri, R., Kerrest, A., Koszul, R., Lemaire, M., Lesur, I., Ma, L., Muller, H., Nicaud, J.M., Nikolski, M., Oztas, S., Ozier-Kalogeropoulos, O., Pellenz, S., Potier, S., Richard, G.F., Straub, M.L., Suleau, A., Swennen, D., Tekaia, F., Wésolowski-Louvel, M., Westhof, E., Wirth, B., Zeniou-Meyer, M., Zivanovic, I., Bolotin-Fukuhara, M., Thierry, A., Bouchier, C., Caudron, B., Scarpelli, C., Gaillardin, C., Weissenbach, J., Wincker, P., Souciet, J.L. (2004) *Genome evolution in yeasts*. Nature 430 (6995): 35-44

Dujon, B. (2006) *Yeast illustrate the molecular mechanisms of eukaryotic genome evolution*. Trends Genet 22 (7): 357-87

Dunlop, P.C., Meyers, G.M., Roon, R.J. (1978) *Characterization of two forms of aspaginase in Saccharomyces cerevisiae*. J Biol Chem 253: 1297-1304

Dunn, B., Levine, R.P. Sherlock, G. (2005) *Microarray karyotyping of commercial wine yeast strains reveals shared, as well as unique, genomic signatures*. BMC Genomics 6 (1):53

Edwards-Ingram L. C., Gent M. E., Hoyle D. C., Hayes A., Stateva L. I., Oliver S. G. (2004) *Comparative Genomic Hybridization Provides New Insights Into the Molecular Taxonomy of the Saccharomyces Sensu Stricto Complex*. Genome Res. 14: 1043-1051

Eisen, M.B., Spellman, P.T., Brown, P.O., Botstein, D. (1998) *Cluster analysis and display of genome-wide expression patterns*. Proc Natl Acad Sci USA 95 (25): 14863-8

Esteve-Zarzoso, B., Fernández-Espinar, M.T., Querol, A. (2000) *Authentication and identification of Saccharomyces cerevisiae "flor" yeast races involved in sherry aging*. Antonie Van Leeuwenhoek 85 (2): 151-8

- Fay, J.C., McCullough, H.L., Sniegowski, P.D., Eisen, M.B. (2004) *Population genetic variation in gene expression is associated with phenotypic variation in Saccharomyces cerevisiae*. Genome Biol 5 (4):26
- Fink, G., Farabaugh, P., Roeder, G., Chaleff, D. (1981) *Transposable elements (ty) in yeast*. Cold Spring Harb Symp Quant Biol. 45 (2): 575-80
- Fleet, G.H. (2003) *Yeast interactions and wine flavour*. Int J Food Microbiol 86 (1-2):11-22
- Garrett, M.D., Zahner, J.E., Cheney, C.M., Novick, P.J. (1994) *GDII encodes a GDP dissociation inhibitor that plays an essential role in the yeast secretory pathway*. EMBO J. 13 (7): 1718-1728
- Gentleman, R.C., Carey, V.J., Bates, D.M., Bolstad, B., Dettling, M., Dudoit, S., Ellis, B., Gautier, L., Ge, Y., Gentry, J., et al. (2004) *Bioconductor: open software development for computational biology and bioinformatics*. Genome Biol. 5:R80
- Goffeau A. (2000) *Four years of post-genomic life with 6000 yeast genes*. FEBS Letters 480:37-41
- Goffeau A., Barrel, B.G., Bussey, H., Davis, R.W., Dujon, B., Feldmann, H., Galibert, F., Hoheisel, J.D., Jacq, C., Johnston, M., Louis, E.J., Mewes, H.W., Murakami, Y., Philippsen, P., Tettelin, H., Oliver, S.G. (1996) *Life with 6000 genes*. Science 274: 563-567
- Goffeau, A., Vassarotti, A. (1991) *The European project for sequencing the yeast genome*. Res Microbiol. 142 (7-8): 901-3
- Goffeau, A., Aert, R., Agostini-Carbone, M.L., Ahmed, A., Aigle, M., Alberghina, L., Albermann, K., Albers, M., Aldea, M., Alexandraki, D., Aljinovic, G., Allen, E., Alt-Mörbe, J., André, B., Andrews, S., Ansorge, W., Antoine, G., Anwar, R., Aparicio, A., Araujo, R., Arino, J., Arnold, F., Arroyo, J., Aviles, E., Backes, U., Baclet, M.C., Badcock, K., Bahr, A., Baladron, V., Ballesta, J.P.G., Bankier, A.T., Banrevi, A., Bargues, M., Baron, L., Barreiros, T., Barrell, B.G., Barthe, C., Barton, A.B., Baur, A., Bécam, A.-M., Becker, A., Becker, I., Beinhauer, J., Benes, V., Benit, P., Berben, G., Bergantino, E.,

- Bergez, P., Berno, A., Bertani, I., Biteau, N., Bjourson, A.J., Blöcker, H., Blugeon, C., Bohn, C., Boles, E., Bolle, P.A., Bolotin-Fukuhara, M., Bordonné, R., Boskovic, J., Bossier, P., Botstein, D., Bou, G., Bowman, S., Boyer, J., Brandt, P., Brandt, T., Brendel, M., Brennan, T., Brinkman, R., *et al.* (1997) *The Yeast genome directory*. Nature 387 (Suppl.): 5-105
- Gojkovic, Z., Knecht, W., Zameitat, E., Warneboldt, J., Coutelis, J.B., Pynyaha, Y., Neuveglise, C., Moller, K., Löffler, M., Piskur, J., (2004) *Horizontal gene transfer promoted evolution of the ability to propagate under anaerobic conditions in yeasts*. Mol Genet Genetics 271 (4): 387-93
- Greig, D., Louis, E.J., Borts, R.H., Travisano, M. (2002) *Hybrid speciation in experimental populations of yeast*. Science 298 (5599):1773-5
- Guijo, S., Mauricio, J.C., Salmon, J.M., Ortega, J.M. (1997) *Determination of the relative ploidy in different Saccharomyces cerevisiae strains used for fermentation and “flor” film ageing of dry sherry-type wines*. Yeast 13 (2): 101-117
- Guan, Y., Dunham, M.J., Troyanskaya, O.G. 2007 *Functional analysis of gene duplication in Saccharomyces cerevisiae*. Genetics 175 (2): 933-43
- Hall, C., Brachat, S., Dietrich, F.S. (2005) *Contribution of horizontal gene transfer to the evolution of the Saccharomyces cerevisiae*. Eukaryot Cell 4 (6): 1102-15
- Hall, D.A., Zhu, H., Zhu, X., Royce, T., Gerstein, M., Snyder, M. (2004) *Regulation of gene expression by a metabolic enzyme*. Science 5695: 482-484
- Haubold, B., Wiehe, T. (2004) *Comparative genomics: methods and applications*. Naturwissenschaften 91(9): 405-21
- Hauser, N.C., Fellenberg, K., Gil, R., Bastuck, S., Hoheisel, J.D., Pérez-Ortín, J.E. (2001) *Whole genome analysis of a wine yeast strain*. Comp Funct Genom 2: 69-79
- Herskowitz, I., (1992) *Fungal physiology. The yeast branches out*. Natures 357 (6375): 190-1

- Hoffman, C.S., Winston, F. (1987) *A ten-minute DNA preparation from yeast efficiently releases autonomous plasmids for transformation of Escherichia coli*. Gene 57 (2-3):267-72
- Infante, J.J., Dombek, K.M., Rebordinos, L., Cantoral, J.M., Young, E.T. (2003) *Genome-wide amplifications caused by chromosomal rearrangements play a major role in adaptive evolution of natural yeast*. Genetics 165:1745-1759
- Jensen, K.F., Bjornberg, O. (1998) *Evolutionary and functional families of dihydroorotate dehydrogenase*. Paths Pyrimidines 6: 20-28
- Jeong, H.S., Backlund, P.S., Chen, H.C., Karavanov, A.A., Crouch, R.J. (2004) *RNase H2 of Saccharomyces cerevisiae is a complex of three proteins*. Nucleic Acids Res 32 (2): 407-414
- Johnston, J.R., Baccari, C., Mortimer, R.K. (2000) *Genotypic characterization of strains of commercial wine yeasts by tetrad analysis*. Res Microbiol 151 (7): 583-90
- Jordan, I.K., MacDonald, K. (1999) *Tempo and mode of Ty evolution in Saccharomyces cerevisiae*. Genetics 151: 1341-1351
- Kallioniemi, A., Kallioniemi, O.P., Sudar, D., Rutovitz, D., Gray, J.W., Waldman, F., Pinkel, D. (1992) *Comparative genomic hybridization for molecular cytogenetic analysis of solid tumors*. Science 258 (5083): 818-21
- Kim, D.Y., Song, W.Y., Yang, Y.Y., Lee, Y. (2001) *The role of PDR13 in tolerance to high copper stress in budding yeast*. FEBS Lett 508 (1): 99-102
- Kim, J.M., Vanguri, S., Boeke, J.D., Gabriel, A., Voytas, D.F. (2007) *Transposable elements and genome organization: a comprehensive survey of retrotransposons revealed by the complete Saccharomyces cerevisiae genome sequence*. Genome Res. 8: 464-478
- Kim, K.W., Kamerud, J.Q., Livingston, D.M., Roon, R.J. (1988) *Aspariganse II of Saccharomyces cerevisiae*. J Biol Chem 263 (24):11948-11988



- Kinzler, K.W., Vogelstein, B. (1996) *Lessons from hereditary colorectal cancer*. Cell 87(2):159-70
- Kobayashi, O., Hayashi, N., Kuroki, R., Sone, H. (1998) *Region of Flo1 proteins responsible for sugar recognition*. J Bacteriol 180 (24): 6503-6510
- Kolakowski, L.F.Jr, Schloesser, M., Cooperman, B.S. (1988) *Cloning, molecular characterization and chromosome localization of the inorganic pyrophosphatase (PPA) gene from S. cerevisiae*. Nucleic Acids Res 16 (22): 10441-52
- Koszul R., Caburet, S., Dujon, B., Fischer, G. (2004) *Eucaryotic genome evolution through the spontaneous duplication of large chromosomal segments*. EMBO J. 23 (1): 234-43
- Koszul R., Dujon, B., Fischer, G., (2006) *Stability of large segmental duplication in the yeast genome*. Genetics 172 (4): 2211-22
- Kupiec, M., Petes, T.D. (1988) *Allelic and ectopic recombination between Ty elements in yeast*. Genetics 119 (3): 549-59
- Kupiec, M., Petes, T.D., *Meiotic recombination between repeated transposable elements in Saccharomyces cerevisiae*. Mol Cell Biol 8 (7): 2942-54
- Kupiec M., Sharan R., Ruppin E. (2007) *Genetic interactions in yeast: is robustness going bust?* Molecular Systems Biology 3 (97)
- Landry, C.R., Lemos, B., Rifkin, S.A., Dickinson, W.J., Hartl, D.L. (2007) *Genetic properties influencing the evolvability of gene expression*. Science 317 (5834):118-21
- Lashkari, D.A., DeRisi, J.L., McCusker, J H., Namath, A.F., Gentile, C., Hwang, S.Y., Brown, P.O., Davis, R.W. (1997) *Yeast microarrays for genome wide parallel genetic and gene expression analysis*. Proc. Natl. Acad. Sci. USA 94: 13057–13062
- Lesage, P., Todeschini, A.L. (2005) *Happy together: the life and times of Ty retrotransposons and their hosts*. Cytogenet Genome Res 110 (1-4): 70-90. Review
- Lewin, B (2004) *Genes VIII*, International Edition, Upper Saddle River, Pearson Prentice Hall

- Liti, G., Louis, E.J. (2005) *Yeast evolution and comparative genomics*. Annu Rev Microbiol 59: 135-53
- Llorente, B., Fairhead, C., Dujon, B. (1999) *Genetic redundancy and gene fusion in the genome of the baker's yeast Saccharomyces cerevisiae: functional characterization of a three-member gene family involved in the thiamine biosynthetic pathway*. Mol Microbiol 32 (6): 1140-1152
- Lockhart, D.J., Winkler, E.A. (2000) *Genomics, gene expression and DNA arrays*. Nature 405 (6788):827-36.
- Lodish, H., Berk, A., Zipursky, S.L., Matasudaira, P., Baltimore, D., Darnell, J. (2001) *Molecular Cell Biology*, 4<sup>th</sup> Ed. New York, W.H. Freeman and Company
- Louis, E.J., Harber, J.E. (1992) *The structure and evolution of subtelomeric Y' repeats in Saccharomyces cerevisiae*. Genetics 131 (3): 559-74
- Louis, E.J., Harber, J.E. (1990) *Mitotic recombination among subtelomeric Y' repeats in Saccharomyces cerevisiae*. Genetics 124 (3): 547-59
- Louis, E.J., Harber, J.E. (1990) *The subtelomeric Y' repeat family in Saccharomyces cerevisiae: an experimental system for repeated sequence evolution*. Genetics 124 (3): 533-45
- Madeira, S.C. & Oliveira, A.L. (2004) *Biclustering algorithms for biological data analysis: a survey*. IEEE/ACM Trans Comput Biol Bioinform 1 (1): 24-45
- Masneuf, I., Hansen, J., Groth, C., Piskur, J., Dubordieu, D. (1998) *New hybrids between Saccharomyces sensu stricto yeast species found among wine and cider production strains*. Appl Environ Microbiol 64 (10): 3887-3892
- McAda, P. & Douglas, M.G. (1982) *A neutral metallo endoprotease involved in the processing of an F1-ATPase subunit precursor in mitochondria*. J Biol Chem 257: 3177-3182

- McCusker, J.H., Clemons, K.V., Stevens, D.A., Davis, R.W. (1994) *Genetic characterization of pathogenic Saccharomyces cerevisiae isolate*. Genetics 136 (4):1261-9
- McCusker, J.H., Clemons, K.V., Stevens, D.A., Davis, R.W. (1994) *Saccharomyces cerevisiae virulence phenotype as determined with CD-1 mice is associated with the ability to grow at 42 degrees C and form pseudohyphae*. Infect Immun 62( 12): 5447-55
- Miklos, I., Varga, T., Nagy, A., Sipiczki, M. (1997) *Genome instability and chromosomal rearrangements in a heterothalic wine yeast*. J Basic Microbiol 37 (5): 345-354
- Moller, K., Olsson, L., Piskur, J. (2001) *Ability for anaerobic growth is not sufficient development of the petite phenotype in Saccharomyces kluyveri*. J Bacteriol 183: 2485-2489
- Moore, J.K., Harber, J.E. (1996) *Capture of retrotransposon DNA at the sites of chromosomal double-strand breaks*. Nature 383 (6601): 644-6
- Moore, J.K., Harber, J.E. (1996) *Cell cycle and genetic requirements of two pathways of nonhomologous end-joining repair of double-strand breaks in Saccharomyces cerevisiae*. Mol Cell Biol 16 (5): 2164-73
- Mortimer, R.K., Schild, D. (1985) *Genetic Map of Saccharomyces cerevisiae*. Microbiol Rev 49 (3): 181-213
- Mortimer, R.K. (2000) *Evolution and variation of the yeast (Saccharomyces) genome*. Genome Res 10 (4):403-9
- Murali, T.M. & Kasif, S. (2003) *Extracting conserved gene expression motifs from gene expression data*. Pac Symp Biocomput 8: 77-88
- Nadal, D., Colomer, B., Pina, B. (1996) *Molecular polymorphism distribution in phenotypically distinct populations of wine yeast strains*. Appl Environ Microbiol 62 (6): 1944-50
- Nagy, M., Lacroute, F., Thomas, D. (1992) *Divergent evolution of pyrimidine biosynthesis between anaerobic and aerobic yeasts*. Proc Natl Acad Sci USA 89: 8966-8970

- Nosaka, K., Kaneko, Y., Nishimura, H., Iwashima, A. (1993) *Isolation and characterization of a thiamine pyrophosphokinase gene, THI80, from Saccharomyces cerevisiae*. J Biol Chem 268 (23): 17440-7
- Oliver, S.G., van der Aart, Q.J., Agostoni-Carbone, M.L., Aigle, M., Alberghina L., Alexandraki, D., Antoine, G., Anwar, R., Ballesta, J.P., Benit, P., et al. (1992) *The complete DNA sequence of yeast chromosome III*. Nature 357 (6373): 38-46
- Ooi, S.L., Shoemaker, D.D., Boeke, J.D. (2003) DNA helicase gene interaction network defined using synthetic lethality analyzed by microarray. Nature Genet 35: 277-286
- Peyroche, A., Paris, S., Jackson, C.L. (1996) *Nucleotide exchange on ARF mediated by yeast Gea1 protein*. Nature 384(6608): 479-81
- Piarroux, R., Millon, L., Bardonnnet, K., Vagner, O., Koenig, H. (1999) *Are live Saccharomyces yeasts harmful to patients?* Lancet 353 (9167): 1851-2
- Piskur, J. & Langaer, R. B. (2004) *Yeast genome sequencing: the power of comparative genomics*. Mol. Microbiol. 53: 381-389
- Pogoreve, M. & Faber, K. (2003) *Purification and characterization of an inverting stereo- and enantioselective sec-alkylsulfatase from the gram-positive bacterium Rhodococcus rubber DSM 44541*. Appl Environ Microbiol 69: 2810-2815
- Pollack, J.R., Perou, C.M., Alizadeh, A.A., Eisen, M.B., Pergamenschikov, A., Williams, C.F., Jeffrey, S.S., Botstein, D., Brown, P.O. (1999) *Genome-wide analysis of DNA copy-number changes using cDNA microarrays*. Nature Genet 23 (1): 41-6
- Pollack, J.R., Perou, C.M., Alizadeh, A.A., Eisen, M.B., Pergamenschikov, A. (1999) *Genome-wide analysis of DNA copy-number changes using cDNA microarrays*. Nat Genet 23 (1): 41-6
- Pretorius, I.S., Van der Westhuizen, T.J. (1991) *The impact of yeast genetics and recombinant DNA technology on the wine industry – a review*. S Afr J Enol Vitic 12: 3-31

- Pretorius, I.S. (2000) *Tailoring wine yeast for the new millennium: novel approaches to the ancient art of winemaking*. Yeast 16 (8): 675-729
- Pryde, F.E., Gorham, H.C., Louis, E.J. (1997) *Chromosome end: all the same under their caps*. Curr Opin Genet Dev 7 (6): 822-8
- Puig, S., Pérez-Ortín, J.E. (2000) *Expression levels and patterns of glycolytic yeast genes during wine fermentation*. Syst Appl Microbiol. 23 (2): 300-3
- Puig, S., Querol, A., Barrio, E., Pérez-Ortín, J.E. (2000) *Mitotic recombination and genetic changes in Saccharomyces cerevisiae during wine fermentation*. Appl Environ Microbiol 66 (5): 2057-61
- Puig, S., Pérez-Ortín, J.E. (2000) *Stress response and expression patterns in wine fermentations of yeast genes induced at the diauxic shift*. Yeast 16 (2): 139-48
- Rachidi, N., Barre, P., Blondin, B. (1999) *Multiple Ty-mediated chromosomal translocations lead to karyotype changes in a wine strain of Saccharomyces cerevisiae*. Mol Gen Genet 261 (4-5):841-50
- Rainieri, S., Zamonelli, C., Kaneko, Y. (2003) *Saccharomyces sensu stricto: Systematics, genetic diversity and evolution*. J Biosc Bioeng 96 (1): 1-9
- Roeder, G.S. & Fink, G.R. (1982) *Construction of yeast strains containing genetically marked transposons*. Fed Proc 41 (10): 2653-5
- Sabate, J., Cano, J., Esteve-Zaragoza, B., Guillamon, J.M. (2002) *Isolation and identification of yeast associated with vineyard and winery by RFLP analysis of ribosomal genes and mitochondrial DNA*. Microbiol Res 157 (4): 267-74
- Saccharomyces* Genome Database. [<http://www.yeastgenome.org/>]
- Schena, M., Shalon, D., Davis, R.W., Brown, P.O. *Quantitative monitoring of gene expression patterns with a complementary DNA microarrays*. Science 270 (5235): 467-70
- Shampay, J., Szostak, J.W., Blackburn, E.H. (1984) *DNA sequences of telomeres maintained in yeast*. Nature 310 (5973): 154-7

- Simon, R.M., Korn, E.L., McShane, L., Radmocher, M.D., Wright, G.W., Zhao, G. (2003) *Class discovery in Design and design of DNA microarray investigation statistics for biology and health*, New York, Springer
- Snijders, A.M., Nowak, N., Segreaves, R., Blackwood, S., Brown, N., Conroy, J., Hamilton, G., Hindle, A.K., Huey, B., Kimura, K., Law, S., Myambo, K., Palmer, J., Ylstra, B., Yue, J.P., Gray, J.W., Jain, A.N., Pinkel, D., Albertson, D.G. (2001) *Assembly of microarrays for genome-wide measurement of DNA copy number*. Nat Genet 29 (3): 263-4
- Spinelli, S.L., Consaul, S.A., Phizicky, E.M. (1997) *A conditional lethal yeast phosphotransferase (tpt1) mutant accumulates tRNAs with a 2'-phosphate and an undermodified base at the splice junction*. RNA 3 (12): 1388-400
- Steinmetz, L.M., Sinha, H., Richards, D.R., Spiegelman, J.I., Oefner, P.J., McCusker, J.H., Davis, R.W. (2002) *Dissecting the architecture of a quantitative trait locus in yeasts*. Nature 416 (6878): 326-30
- Swiegers, J.H., Dippenaar, N., Pretorius, I.S., Bauer, F.F. (2001) *Carnitine-dependent metabolic activities in Saccharomyces cerevisiae: three carnitine acetyltransferases are essential in a carnitine-dependent strain*. Yeast 18 (7):585-95
- Takahashi, H., McCaffery, J.M., Irizarry, R.A., Boeke, J.D. (2006) *Nucleocytoplasmic acetyl-coenzyme A synthetase is required for histone acetylation and global transcription*. Mol Cell 23: 207-217
- Tarun, S.Z., Wells, S.E., Deardorff, J.A., Sachs, A.B. (1997) *Translation initiation factor eIF4G mediates in vitro poly(A) tail-dependent translation*. PNAS USA 94: 9046-9051
- Teunissen, A.W.R.H. & Steensma, H.Y. (1995) *Review: the dominant flocculation genes in Saccharomyces cerevisiae constitute a new subtelomeric gene family*. Yeast 11: 735-745
- Tusher, V.G., Tibshirani, R., Chu, G. (2001) *Significance analysis of microarrays applied to the ionizing radiation response*. PNAS USA 98 (9): 5116-21
- Van Burik, J.A., Magee, P.T. (2001) *Aspects of fungal pathogenesis in humans*. Annu Rev Microbiol 55: 743-72

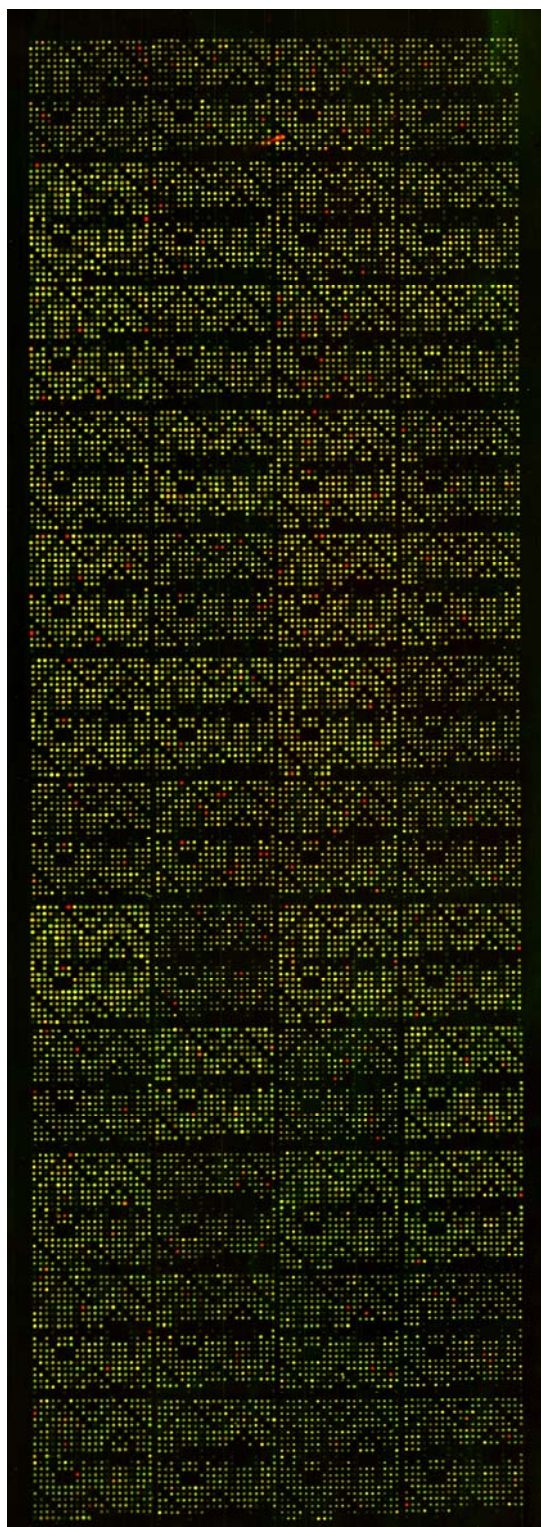
- Wang, P., Kim, Y., Pollack, J., Narasimhan, B., Tibshirani, R. (2005) *A method for calling gains and losses in array CGH data*. Biostatistics 6 (1): 45-58
- White, J.M. & Rose, M.D. (2001) *Yeast mating: Getting close to membrane merger*. Current Biology 11: 16-20
- Winzeler, E.A., Shoemaker, D.D., Astromoff, A., Liang, H., Anderson, K.M., Andre, B., Bangham, R., et al. (1999) *Functional characterization of the S. cerevisiae genome by gene deletion and parallel analysis*. Science 285 (5429): 901-6
- Winkler, E.A., Castillo-Davis, C.I., Oshiro, G., Liang, D., Richards, D.R., Zhou, Y., Hartl, D.L. (2003) *Genetic diversity in yeast assessed with whole-genome oligonucleotide arrays*. Genetics 163 (1): 79-89
- Witte, C., Jensen, R.E., Yaffe, M.P., Schatz, G. (1988) *MAS1, a gene essential for yeast mitochondrial assembly, encodes a subunit of the mitochondrial processing protease*. EMBO J. 7 (5): 1439-1988
- Wolfe, K.H., Shields, D.C. (1997) *Molecular evidence for an ancient duplication of the entire yeast genome*. Nature 387: 708-713
- Xie, W., Gai, X., Zhu, Y., Zappulla, D.C., Sternglanz, R., Voytas, D.F. (2001) *Targeting of the yeast Ty retrotransposon to silent chromatin mediated by interactions between integrase and Sir4p*. Mol cell Biol 21 (19): 6606-14
- Yamada, M., Hayatsu, N., Matsuura, A., Ishikawa, F. (1998) *Y'-Help1, a DNA Helicase encoded by the yeast subtelomeric Y' element, is induced in survivors defective for telomerase*. J Biol Chem 273 (50): 33360-33366
- Yang, M., Jensen, R.E., Yaffe, M.P., Oppliger, W., Schatz, G. (1988) *Import of proteins into yeast mitochondria: the purified matrix processing protease contains two subunits which are encoded by the nuclear MAS1 and MAS2 genes*. EMBO J. 7 (12): 3857-3862



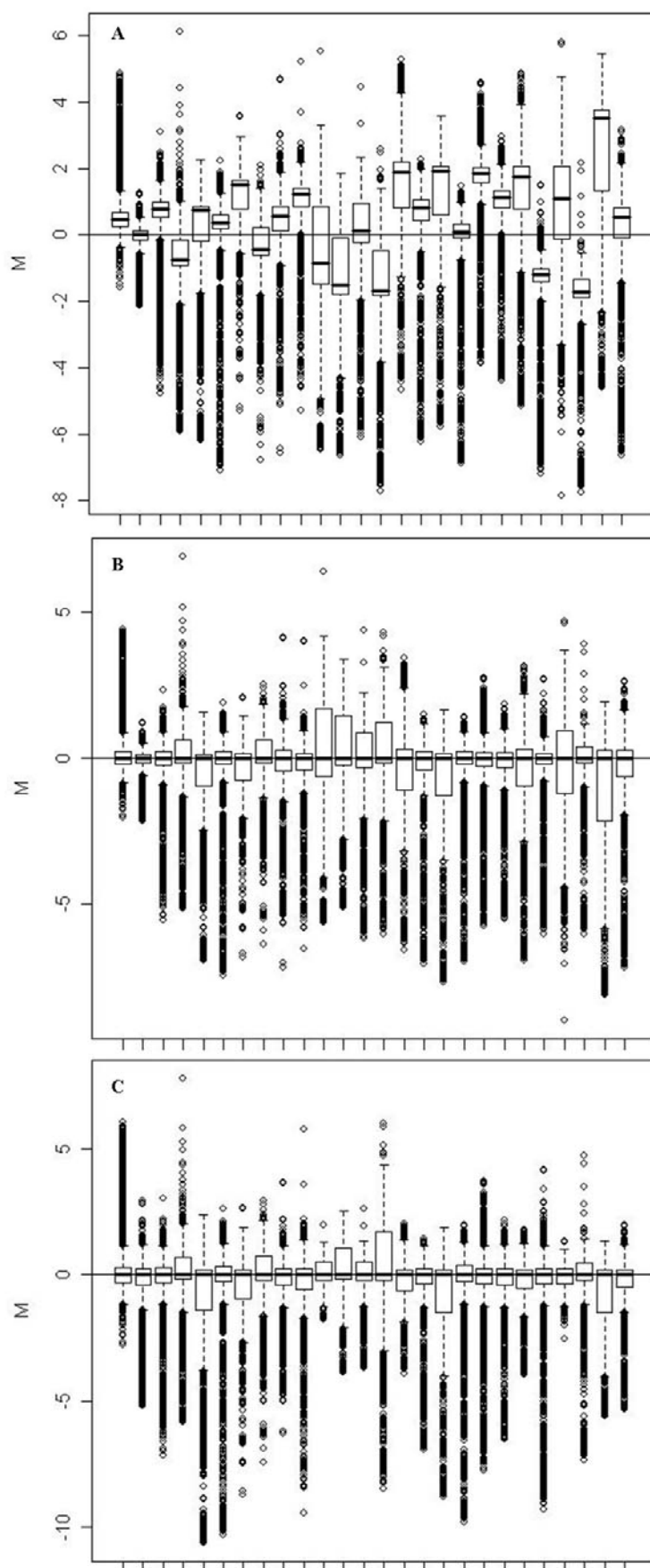


## **APÊNDICES**

**1. Slide de aCGH**



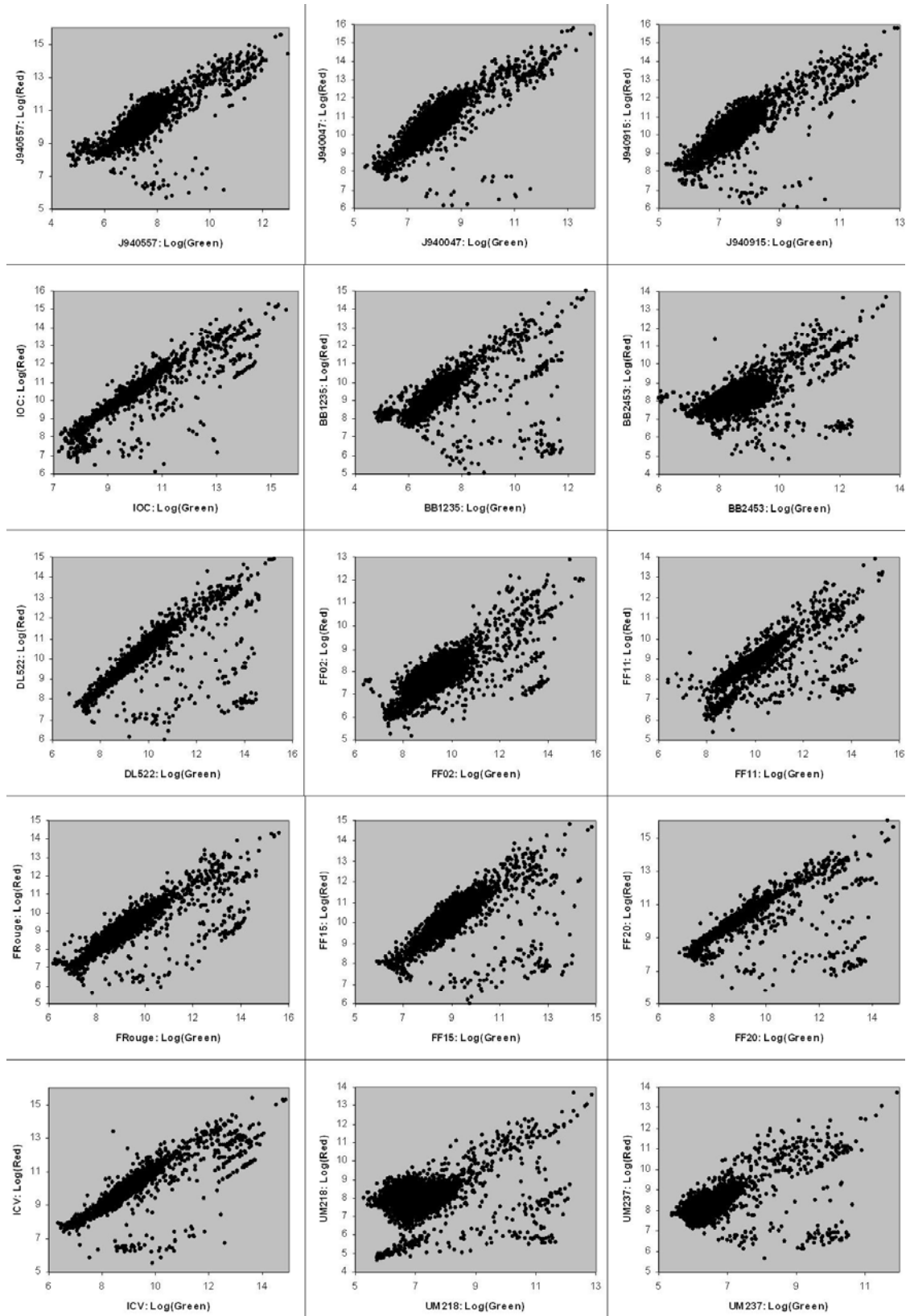
## 2. Gráficos de bigodes

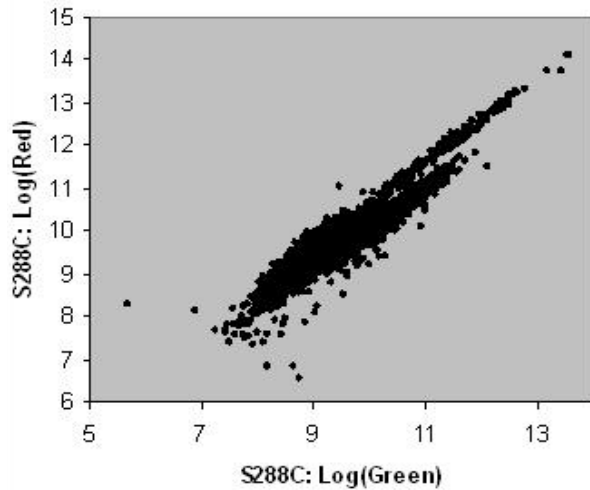


### Representações gráficas da distribuição da intensidade dos sinais de cada uma das cores num slide de aCGH.

Cada caixa representa 50% dos sinais mais intensos, sendo a linha horizontal, a negrito, a mediana destes e estando os *outliers* representados por círculos depois das referidas caixas de bigodes. Uma amostra está representada por duas caixas sucessivas, a primeira correspondendo ao sinal vermelho e a segunda ao sinal verde. O painel A representa a comparação de alguns slides não normalizados, onde se verificou uma variação de distribuição de intensidade global do sinal afectada pela cor. No painel B estão representados os mesmos slides normalizados pela mediana de cada um (centrados) e no painel C, os mesmos slides normalizados pela mediana de cada slide e normalizados uns relativamente aos outros pelo método *Scale*.

### 3. Scatter Plots



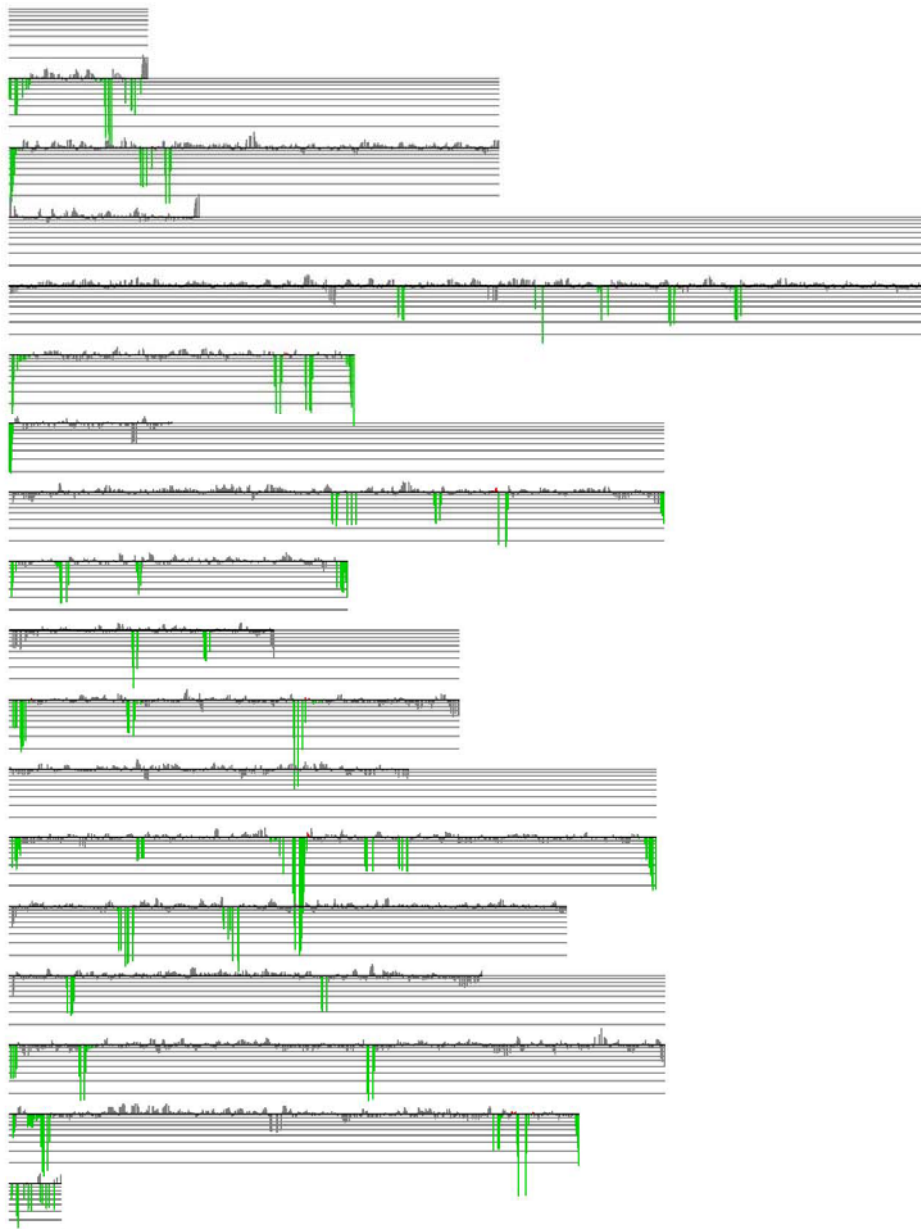


**Representações gráficas do logaritmo da intensidade de uma cor em função do logaritmo da intensidade da outra cor.**

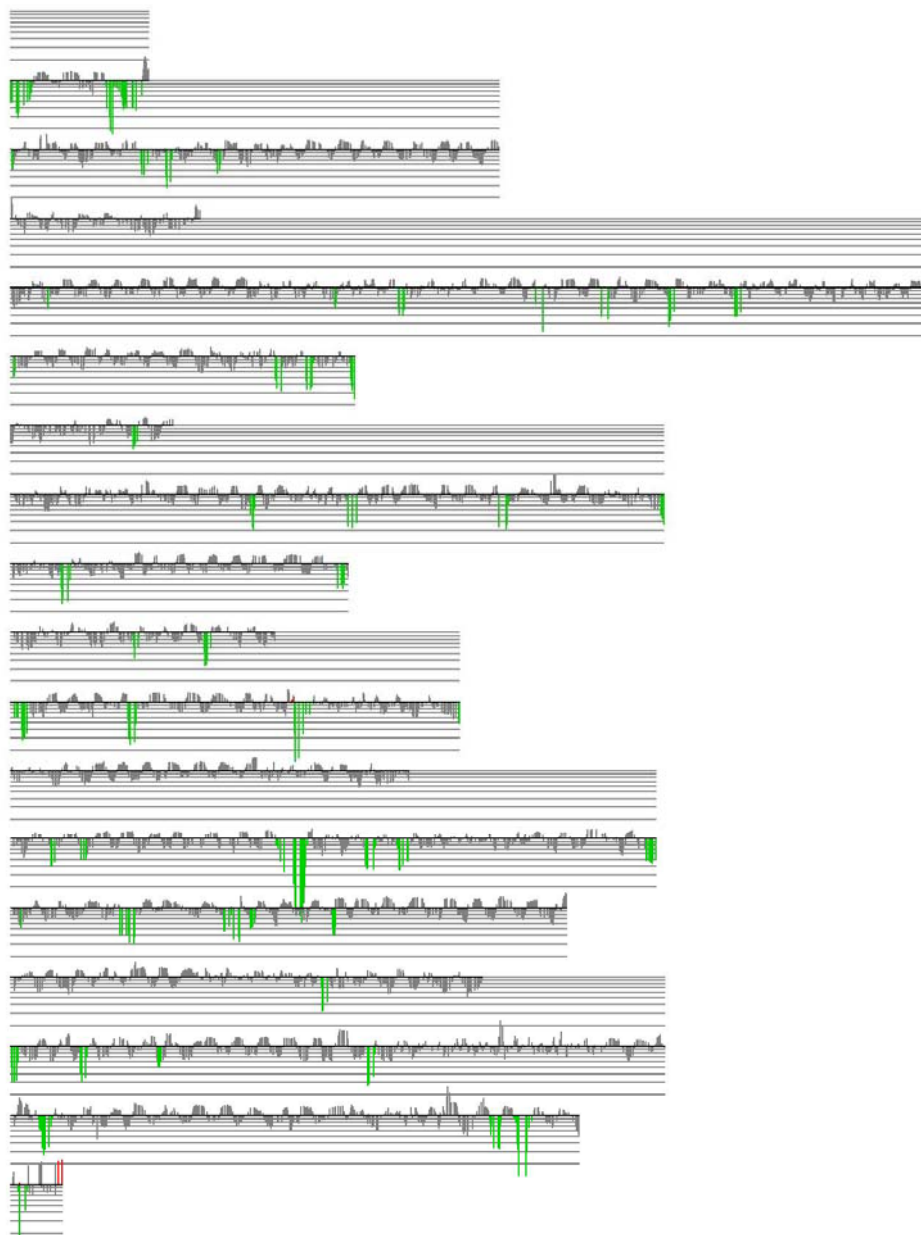
Cada gráfico representa o logaritmo da intensidade de uma cor *versus* o logaritmo da intensidade da outra cor. A maior parte dos sinais localizam-se na nuvem central representando todos os pontos onde ambas as cores têm a mesma intensidade, razão 1:1. Os restantes pontos fora da nuvem central representam genes cujo valor da razão entre amostra e referência é diferente de 1:1, significando diferente número de cópias. Note-se que no gráfico de *S. cerevisiae* S288C, nesta página (#34), todos os sinais estão centrados nesta nuvem de razão 1:1, tal como esperado pois representa uma hibridação *self-self*.

#### 4. Cariótipos das estirpes autóctones de *S. cerevisiae*

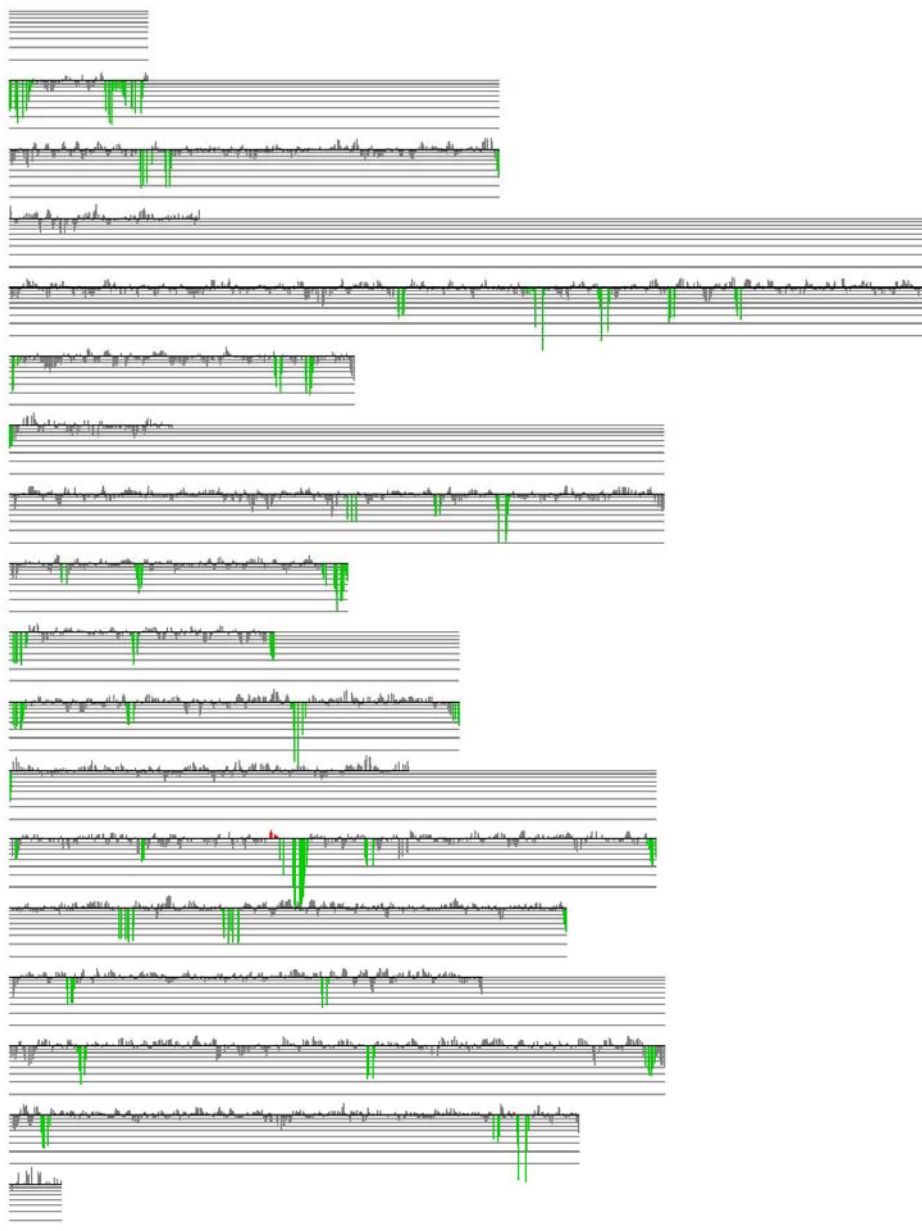
BB1235



BB2453

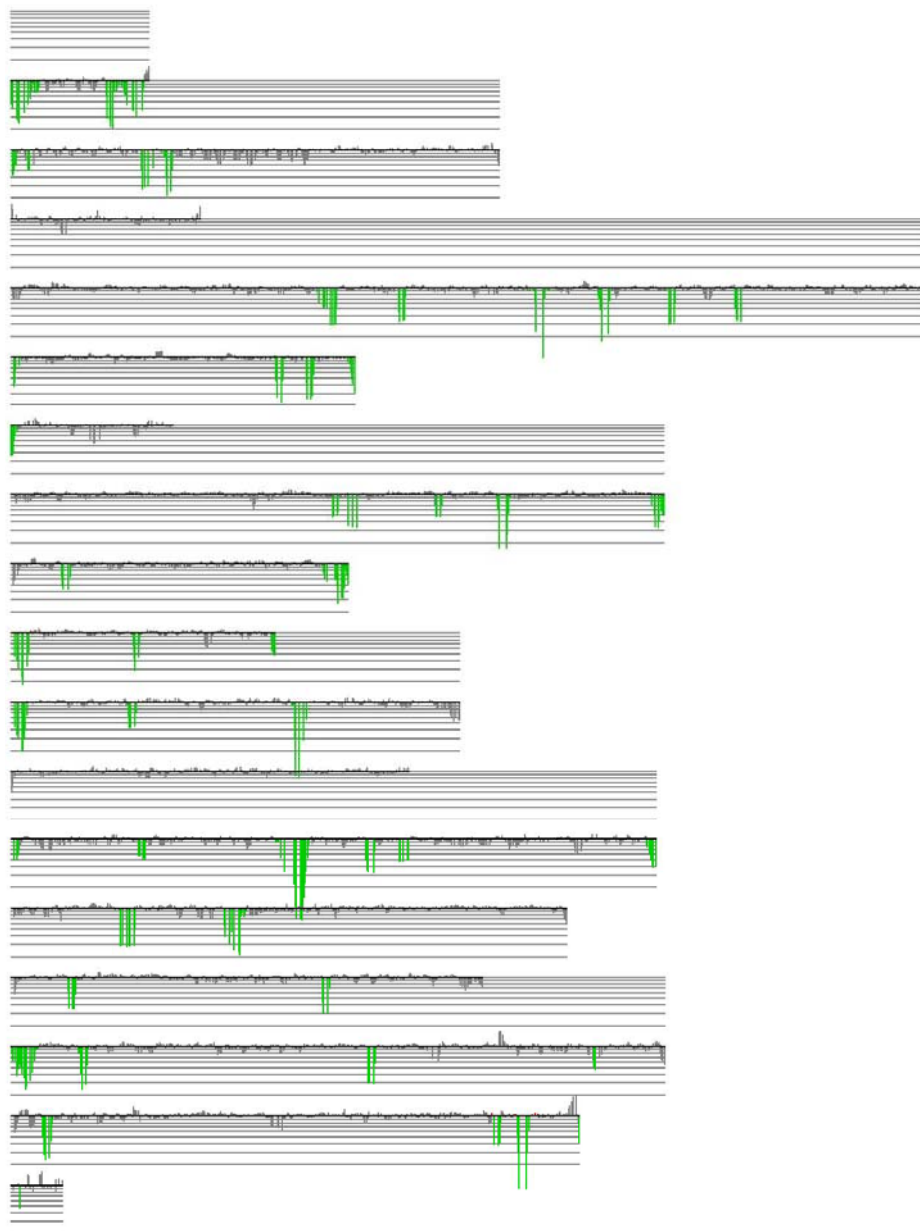


06L3FF02

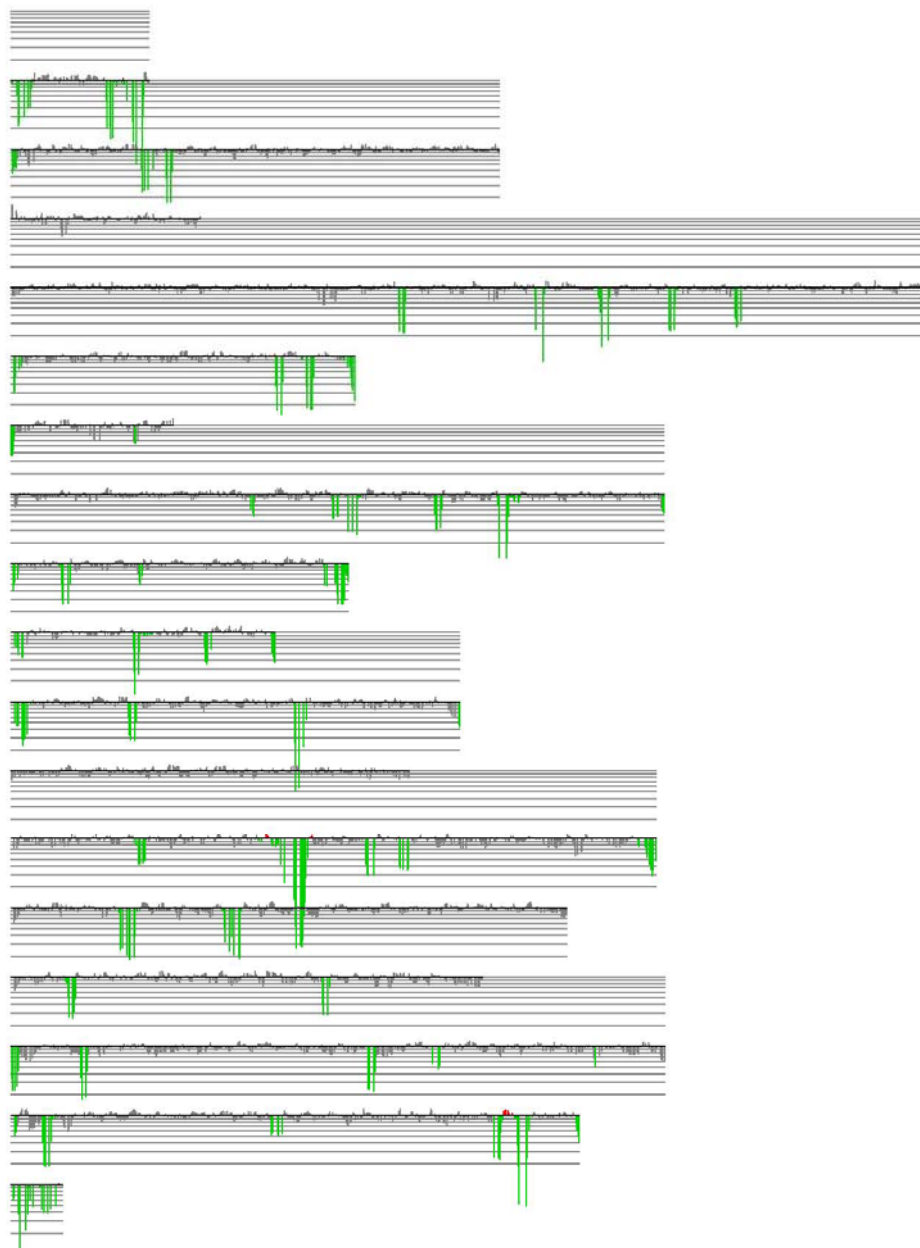




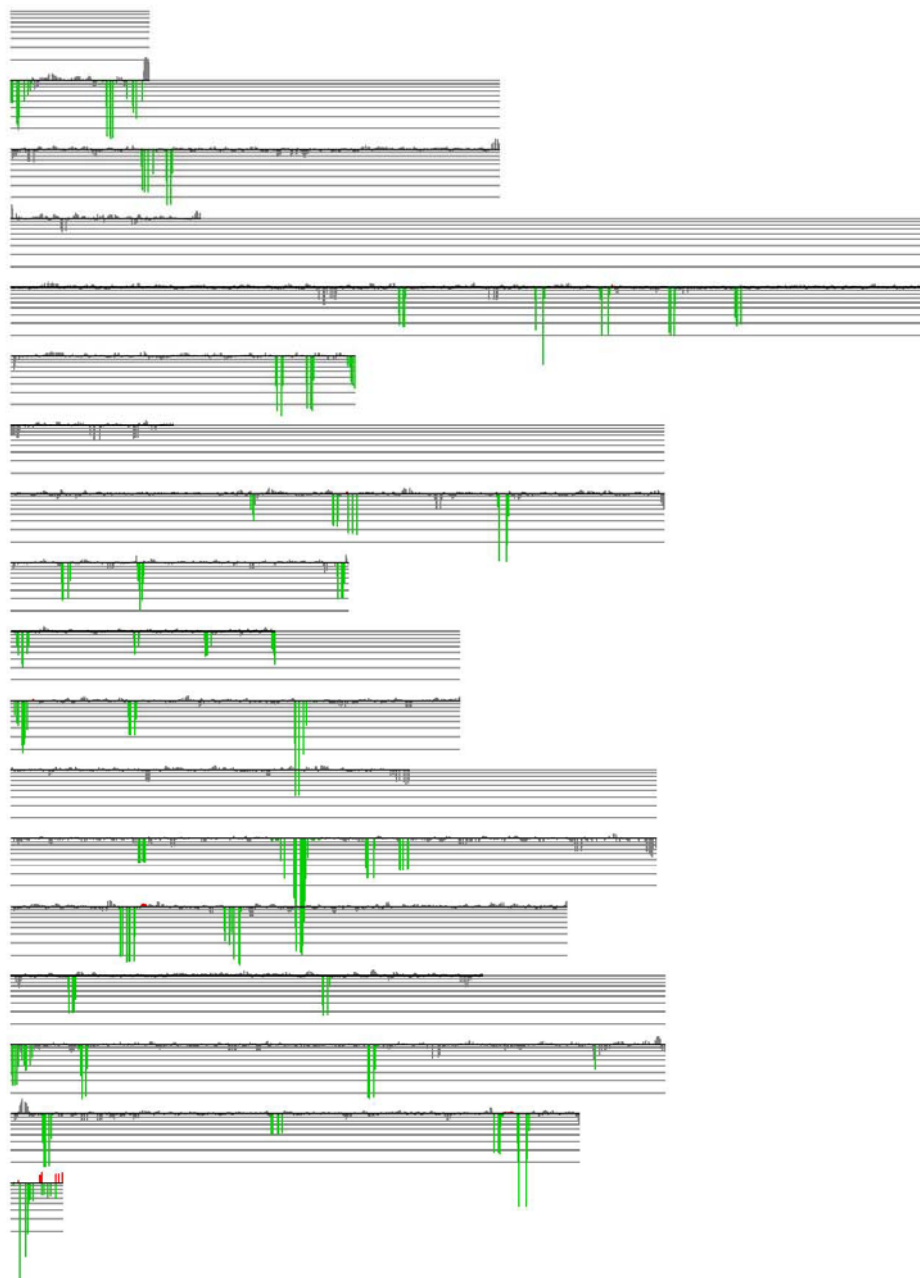
06L1FF11



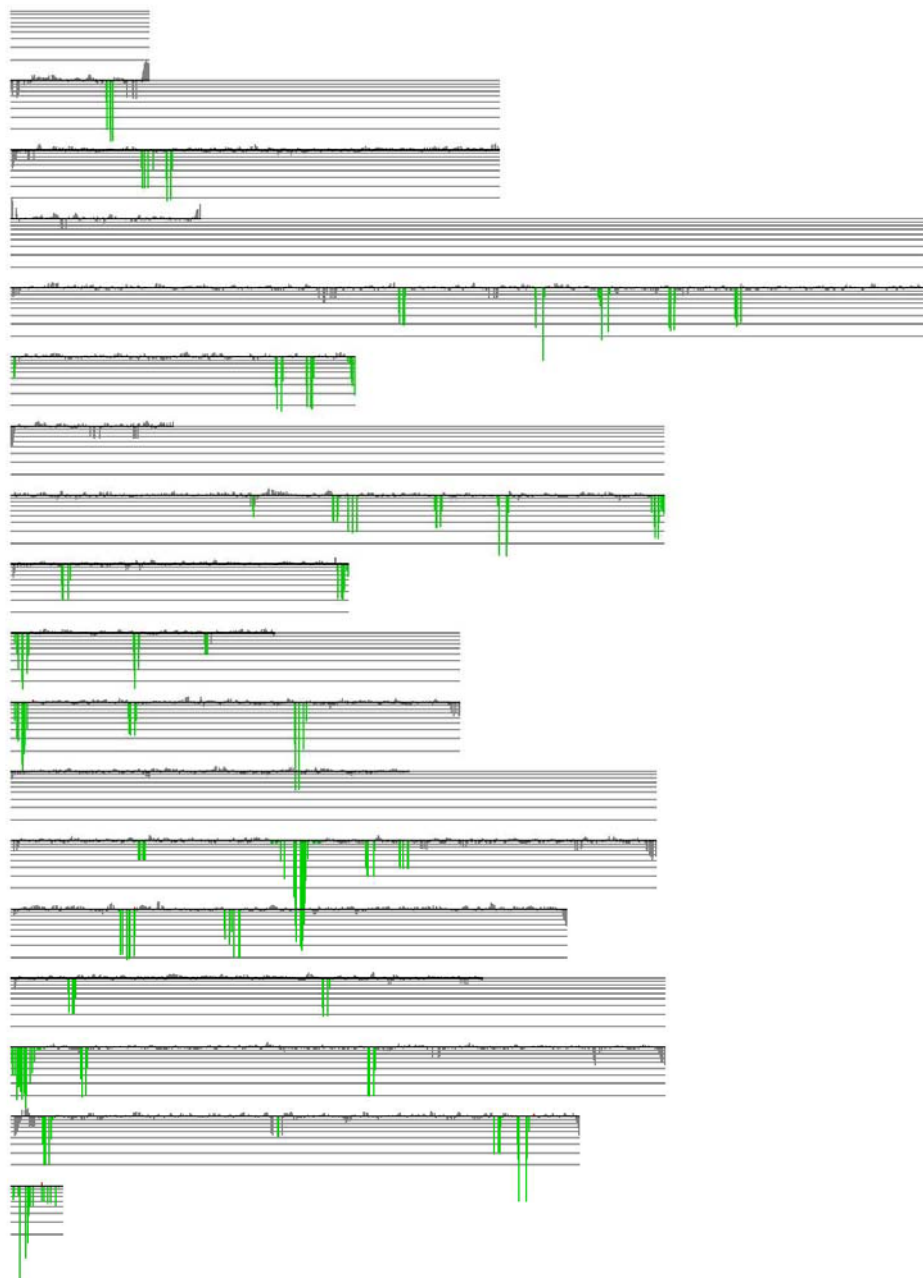
06L3FF15



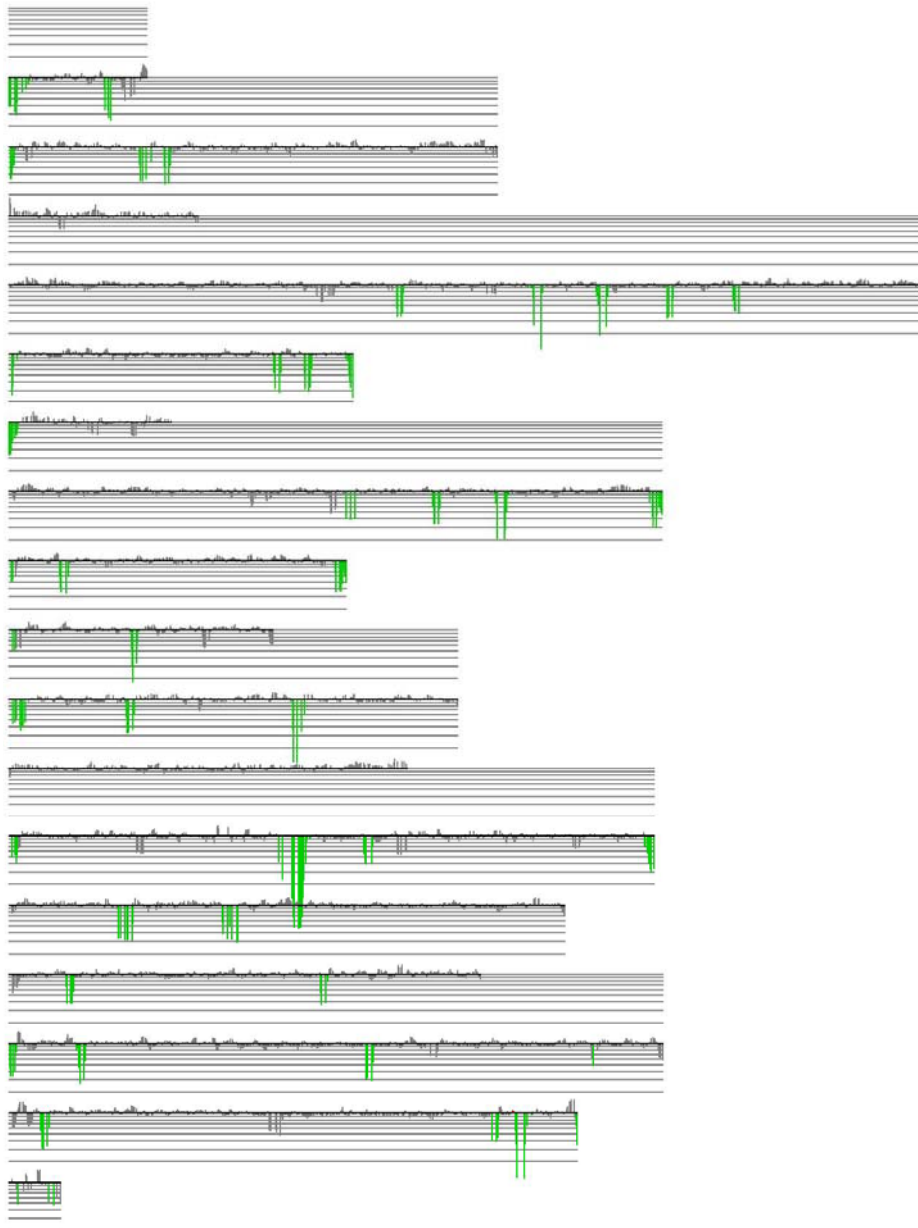
06L6FF20



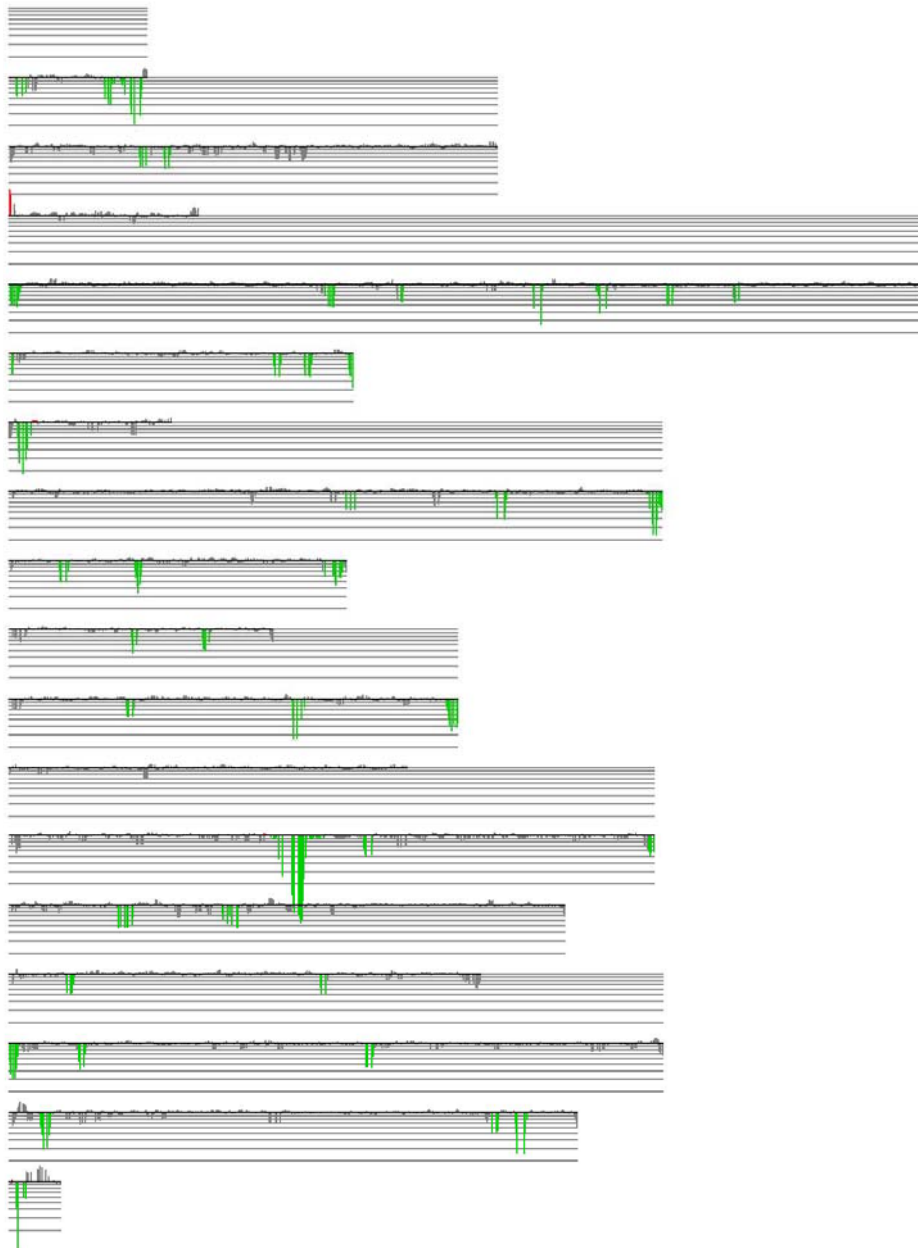
Davis Lalvin 522



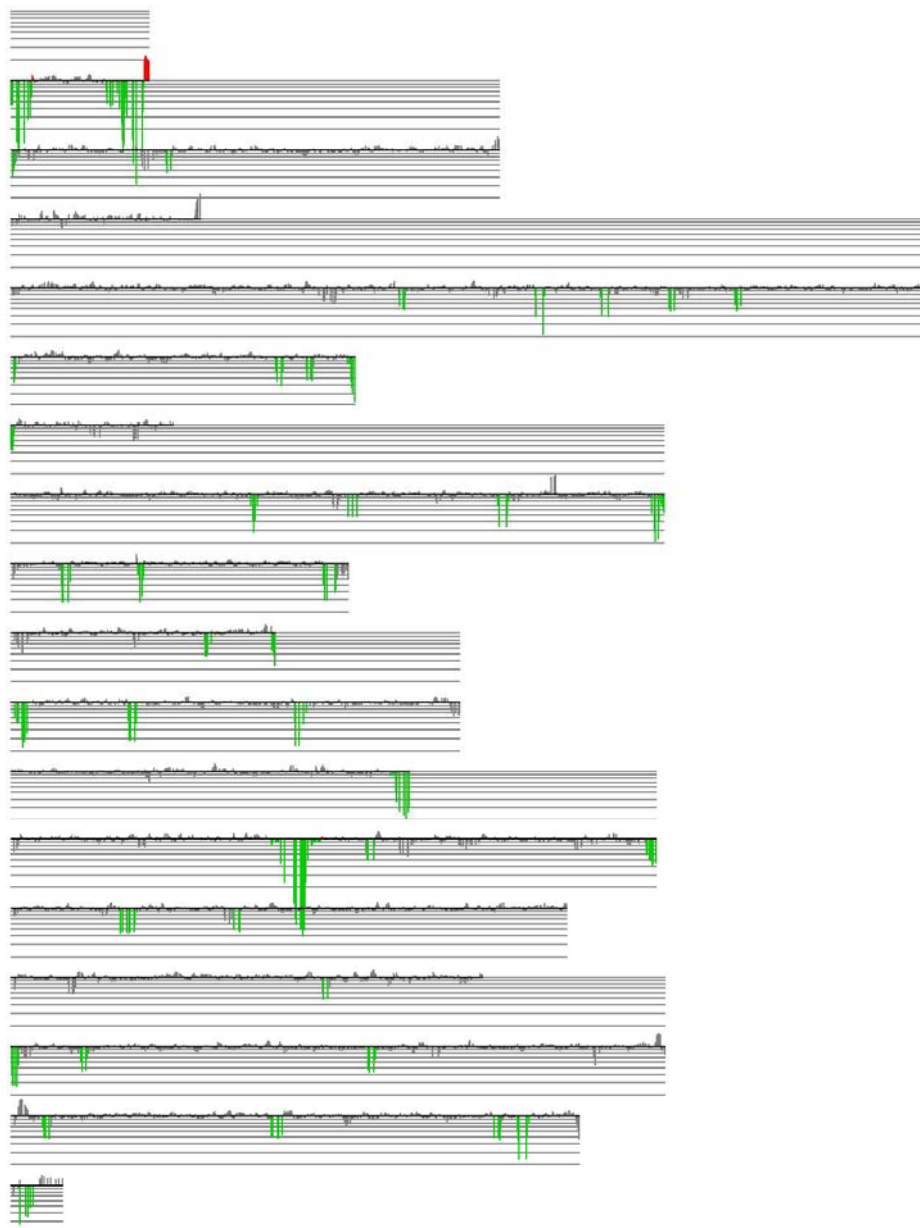
## AEB Fermol Rouge



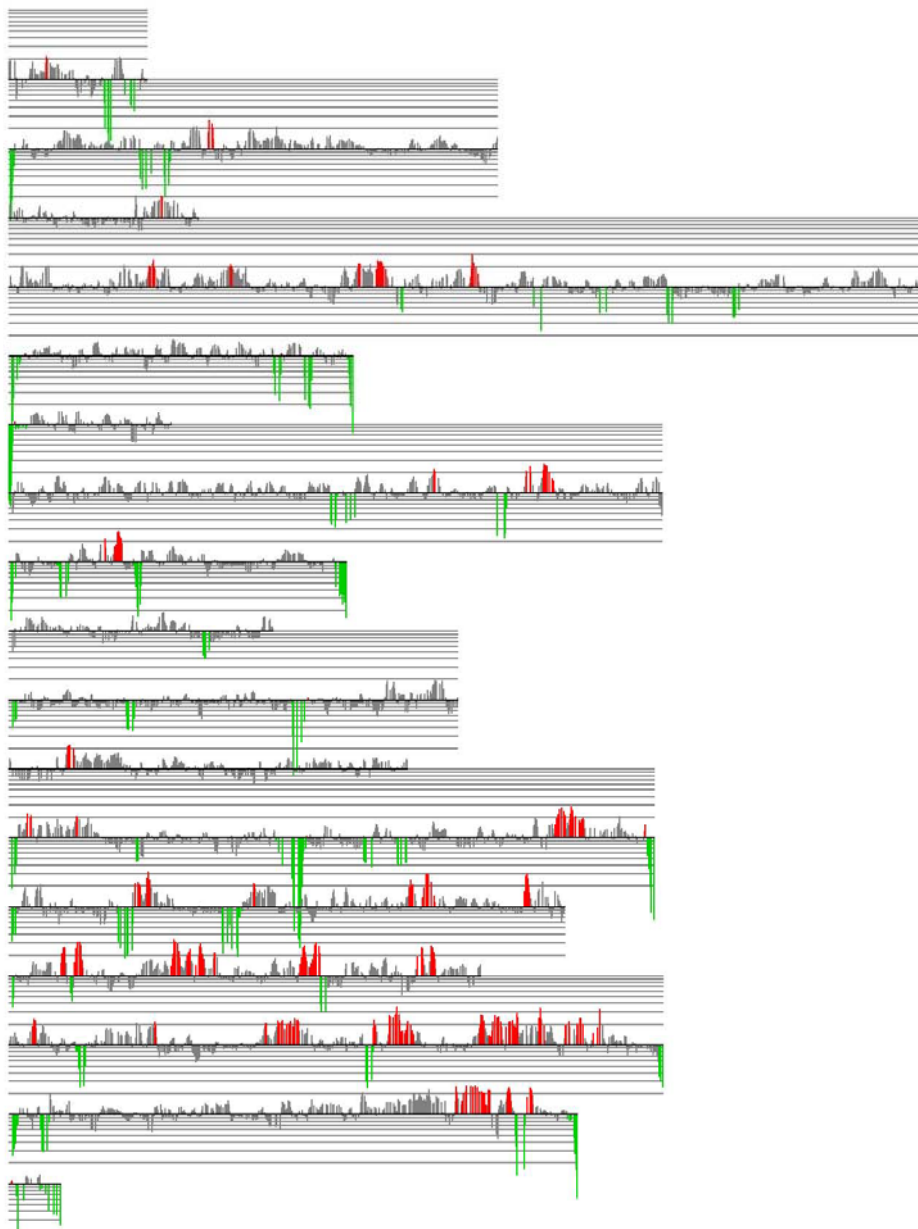
IOC 18



ICV D254

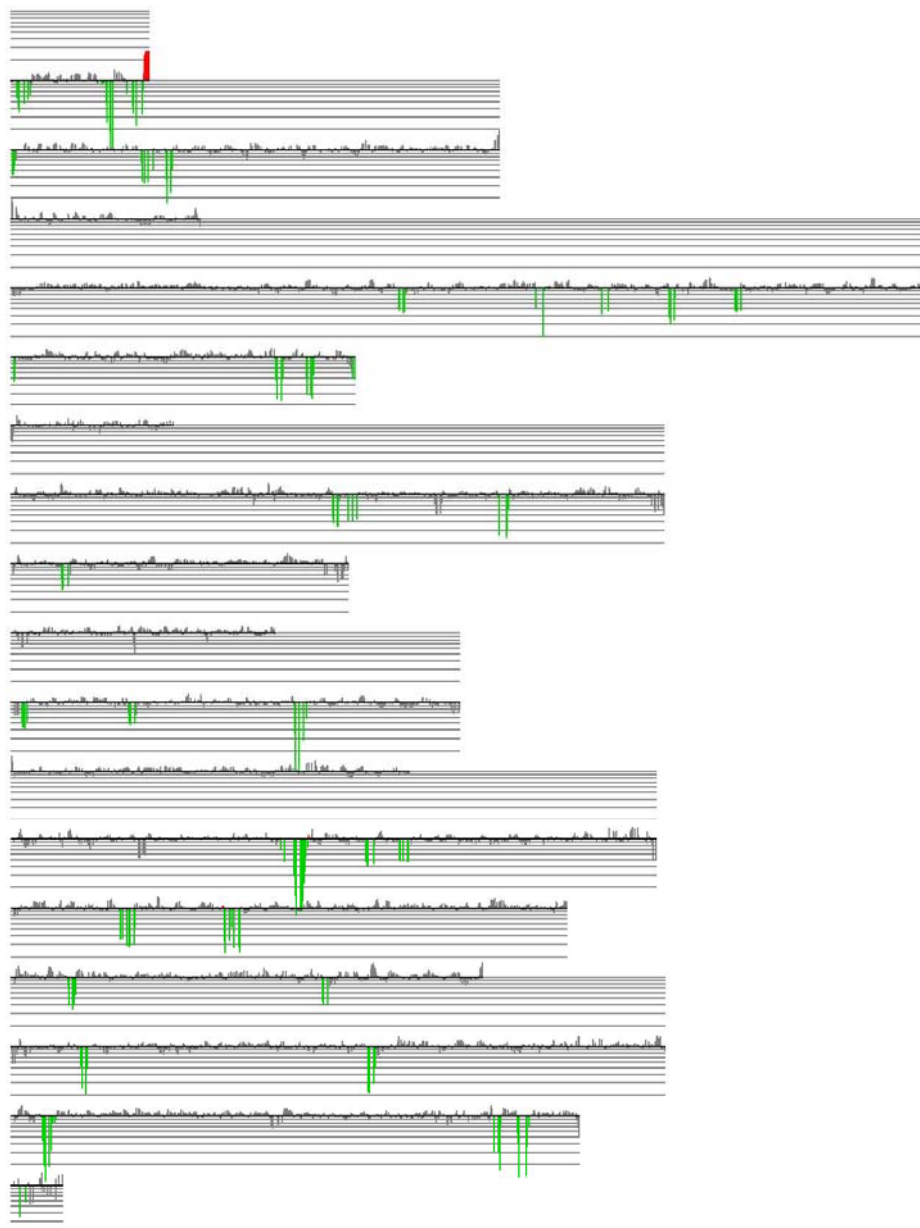


UM218

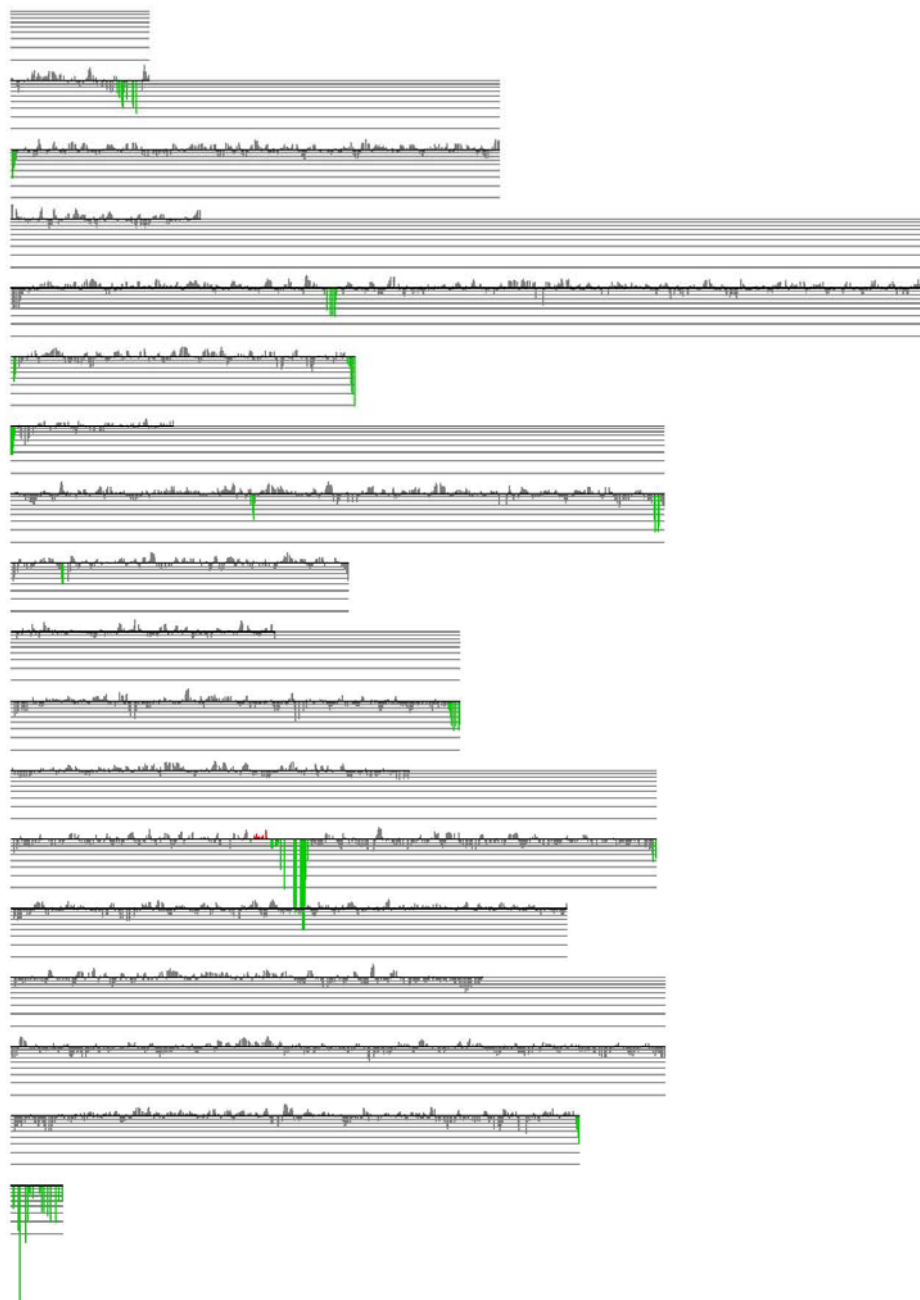




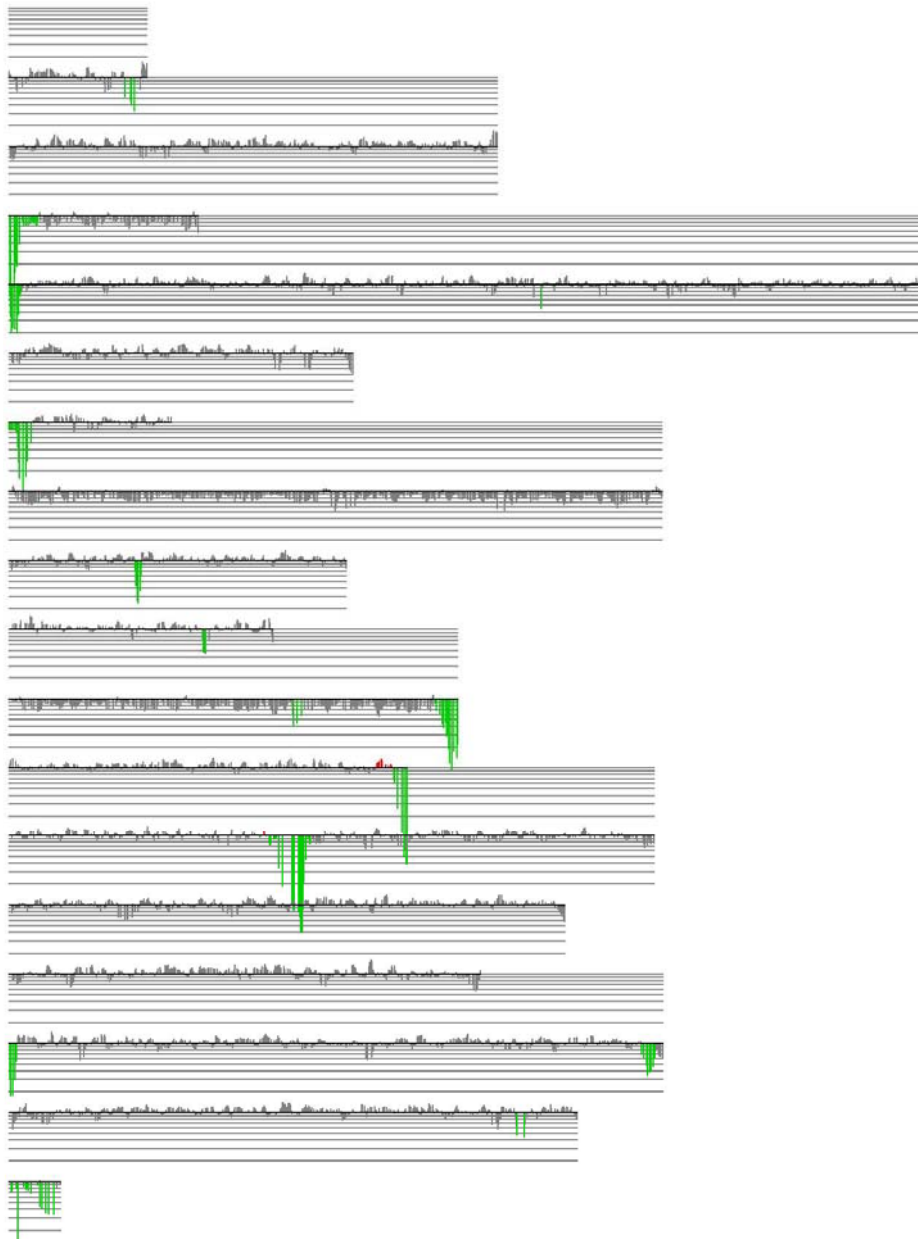
UM237



J940047



J940557



J940915

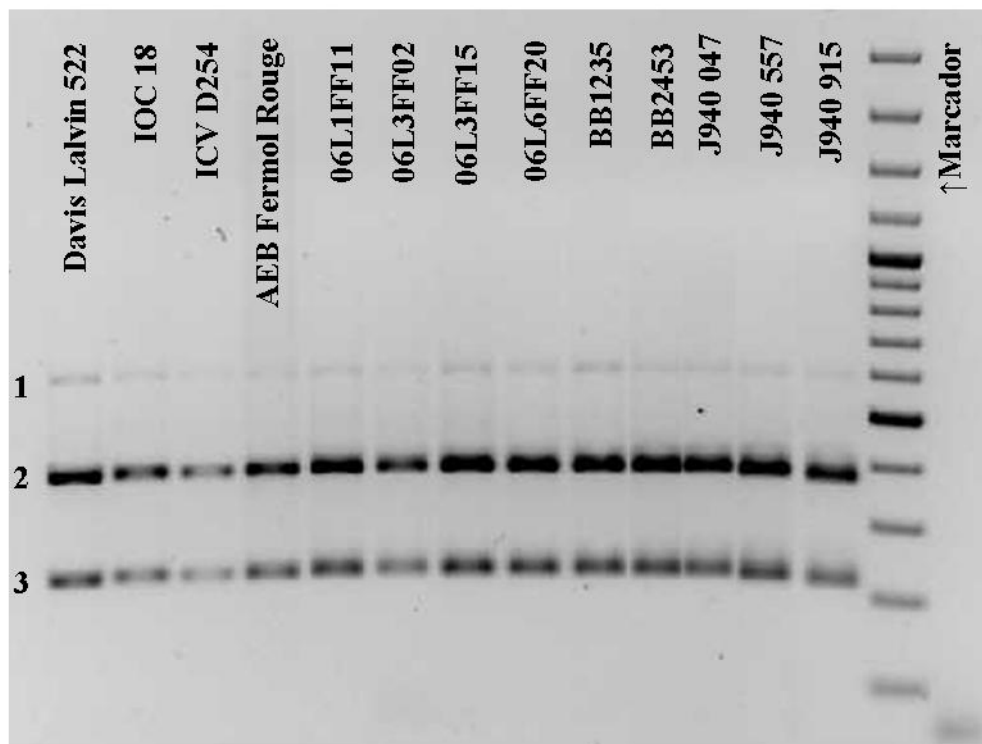






**ANEXOS**

# 1. Perfis da digestão do gene *MET2* com EcoRI

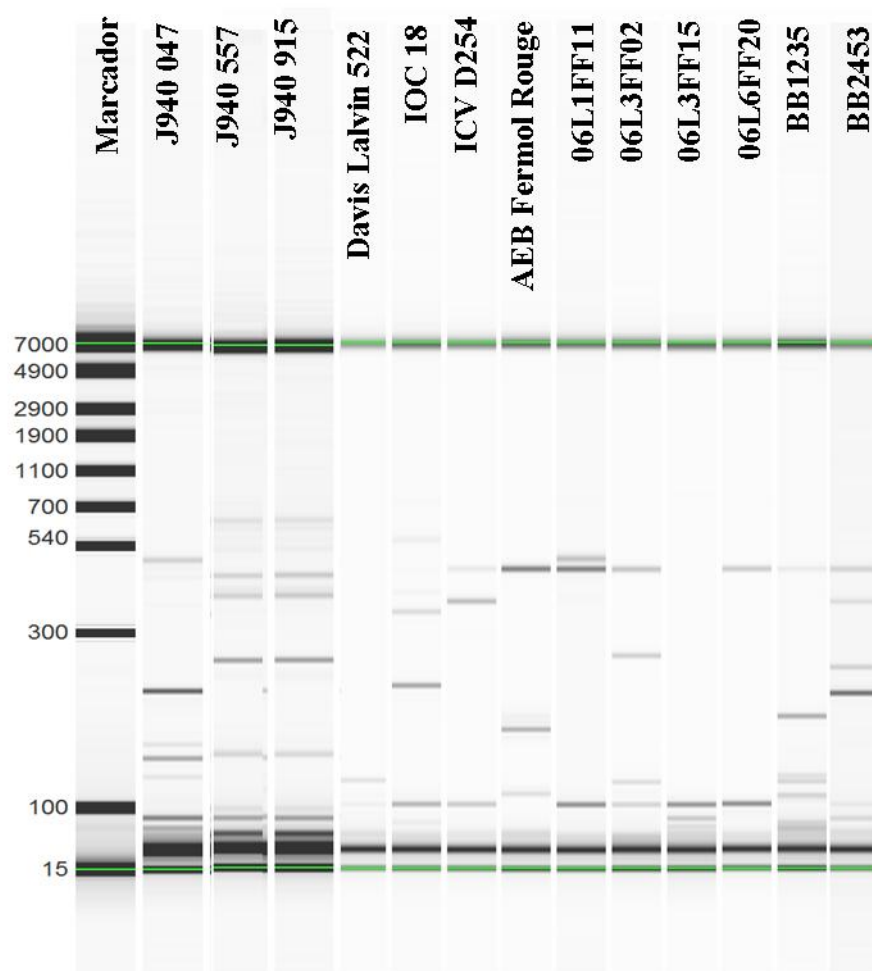


**Perfis da digestão dos produtos de PCR do locus *MET2* com EcoRI [Unidade de Genómica – Biocant].**

O locus *MET2* é clivado pela enzima de restrição EcoRI apenas em *Saccharomyces cerevisiae*. A banda 1 corresponde aos produtos de PCR obtidos para amplificação do gene *MET2*. As bandas 2 e 3 são os fragmentos da digestão com EcoRI dos produtos de PCR anteriormente obtidos (Demuyter C. *et al.*, 2004).



## 2. Perfis Inter- $\delta$



Perfis das regiões inter- $\delta$  das estirpes autóctones *S. cerevisiae*. [Unidade de Genómica – Biocant] (Demuyter *et al.*, 2004)



**ÍNDICE DE TABELAS**

Tabela 1. Características de células <i>S. cerevisiae</i> haploides e diploides.....	6
Tabela 2. Características de <i>mating</i> dos diferentes tipos de células.....	6
Tabela 3. Características fundamentais do genoma de leveduras. ....	8
Tabela 4. Resumo de algumas características dos diferentes tipos de arrays impressos para aCGH.....	19
Tabela 5. Características dos vinhos e estilos de fermentação mais adequados às estirpes de <i>S. cerevisiae</i> industriais analisadas neste estudo .....	32
Tabela 6. Exemplos de genes muito variáveis nas estirpes <i>S. cerevisiae</i> autóctones deste estudo.....	39
Tabela 7. Número total de ORFs detectadas em diferente número de cópias em cada estirpe <i>S. cerevisiae</i> autóctone relativamente a S288C.....	43
Tabela 8. Exemplo de genes apenas em inferior número de cópias em J940557/J940915 comparativamente com J940047. ....	45
Tabela 9. Exemplos de diferentes genes das mesmas famílias com alterações unicamente verificadas na estirpe ICV ou IOC. ....	46
Tabela 10. Nomes de genes e respectivas funções das ORFs obtidas no <i>bi-clustering</i> . ....	53
Tabela 11. Características da assinatura de leveduras vínicas [Dunn <i>et al.</i> (2005)] presentes nas leveduras autóctones deste estudo.....	55
Tabela 12. Assinatura de leveduras vínicas, cromossoma I .....	56
Tabela 13. Assinatura de leveduras vínicas, cromossoma II.....	56
Tabela 14. Assinatura de leveduras vínicas, cromossoma X.....	57
Tabela 15. Assinatura de leveduras vínicas, cromossoma XIII.....	58

## **ÍNDICE DE FIGURAS**

Figura 1. Representação esquemática do ciclo de vida de <i>S. cerevisiae</i> .	7
Figura 2. Constituição dos centrómeros de <i>S. cerevisiae</i> .	9
Figura 3. Representação esquemática de mecanismo de <i>frameshift</i> responsável pelas duas proteínas produzidas a partir dos elementos <i>Ty</i> .	10
Figura 4. Modelo da <i>casset</i> responsável pela troca de tipo sexual ( <i>mating type</i> ).	12
Figura 5. Comparação entre CGH e aCGH.	18
Figura 6. Sistema de marcação fluorescente ULS pela formação de uma ligação coordenada através do Azoto (N7) da Guanina.	24
Figura 7. Esquema da matriz impressa nos slides utilizados em aCGH.	25
Figura 8. Exemplos de "maus spots" obtidos pontualmente nos arrays de aCGH.	29
Figura 9. Gráficos MA para uma estirpe autóctone <i>S. cerevisiae</i> representativa de nichos ecológicos diferentes.	31
Figura 10. Agrupamento hierárquico das 15 estirpes autóctones de <i>S. cerevisiae</i> analisadas juntamente com a estirpe de referência <i>S. cerevisiae</i> S288C.	33
Figura 11. Gráfico consenso das alterações cromossómicas verificadas nas estirpes autóctones de <i>S. cerevisiae</i> analisadas.	35
Figura 12. Distribuição das ORFs com alteração do número de cópias em pelo menos 11 das 15 estirpes autóctones de <i>S. cerevisiae</i> por processo biológico.	37
Figura 13. Mapa da localização relativa de algumas ORFs cujo número de cópias é diferente relativamente à referência <i>S. cerevisiae</i> S288C em todas as estirpes autóctones.	41
Figura 14. Representação gráfica das ORFs onde se verificou alteração no número de cópias em J940557 e J940915 como subgrupo, relativamente a J940047. Distribuição por categoria de função molecular.	44
Figura 15. Comparação da percentagem de ORFs da mesma categoria que em UM218 sofreram aumento e diminuição do número de cópias.	47
Figura 16. Distribuição dos genes do subgrupo R2 relativamente às restantes estirpes de <i>S. cerevisiae</i> para análise SAM.	48
Figura 17. Representação gráfica dos genes significativamente diferentes nas estirpes autóctones de R2 relativamente às restantes estirpes (FDR = 0).	49
Figura 18. Representação gráfica dos genes significativamente diferentes nas estirpes autóctones de UM relativamente às restantes estirpes (FDR = 0).	50

Figura 19. Representação gráfica do resultado do <i>bi-clustering</i> (xMotifs) de amostras e genes. ....	52
-----------------------------------------------------------------------------------------------------------	----